## (19) World Intellectual Property Organization

International Bureau



# 

(43) International Publication Date 12 August 2004 (12.08.2004)

### (10) International Publication Number WO 2004/067743 A1

(51) International Patent Classification7: 15/63, C07H 21/04, A01H 5/00

C12N 15/11,

(21) International Application Number:

PCT/US2003/041269

(22) International Filing Date:

24 December 2003 (24.12.2003)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/441,377	21 January 2003 (21.01.2003)	US
60/441,381	21 January 2003 (21.01.2003)	US
60/441,392	21 January 2003 (21.01.2003)	US
607441,405	21 January 2003 (21.01.2003)	US
60/441,447	21 January 2003 (21.01.2003)	US
60/441,502	21 January 2003 (21.01.2003)	US
10/609,019	26 June 2003 (26.06.2003)	US

- (71) Applicants (for all designated States except US): TRANS-GENRX, INC. [US/US]; 1755 Wittington Place Drive, Suite 300, Dallas, TX 75234 (US). THE BOARD OF SUPERVISORS OF LOUISIANA STATE UNIVER-SITY AND AGRICULTURAL AND MECHANICAL COLLEGE [US/US]; P.O. Box 25055, Baton Rouge, LA 70894 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): COOPER, Richard,

K. [US/US]; 111 Pecan Meadow Drive, Baton Rouge, LA 70810 (US). FIORETTI, William, C. [US/US]; 2225 Lakeridge Road Drive, Grapevine, TX 76051 (US). CADD, Gary, G. [US/US]; 501 Turner Road, Apartment 1111, Grapevine, TX 76051 (US).

- (74) Agent: PRATT, John, S.; Kilpatrick Stockton LLP, Suite 2800, 1100 Peachtree Street, Atlanta, GA 30309 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN. MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

#### Published:

with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: VACCINE PRODUCTION USING TRANSPOSON BASED VECTORS

(57) Abstract: The present invention provides novel methods of making proteins, multimeric proteins and antibodies using transposon-based vectors in transgenic individuals that are desirable in producing vaccines. The present invention further provides a method for providing protection against disease by vaccinating an individual with the proteins, multimeric proteins and/or antibodies made in the transgenic individuals of the present invention.

		4
	•	,
		ř

5

15

# 10 VACCINE PRODUCTION USING TRANSPOSON BASED VECTORS

The U.S. Government has certain rights in this invention. The development of this invention was partially funded by the United States Government under a HATCH grant from the United States Department of Agriculture, partially funded by the United States Government with Formula 1433 funds from the United States Department of Agriculture and partially funded by the United States Government under contract DAAD 19-02016 awarded by the Army.

### 20- FIELD OF THE INVENTION

The present invention relates generally to production of proteins, polypeptides and peptides in a transgenic individual, wherein genes encoding the proteins, polypeptides and peptides are operably-linked to signal sequences, or portions of signal sequences. The proteins, polypeptides and peptides are useful as vaccines.

25

30

### BACKGROUND OF THE INVENTION

Diseases caused by numerous biological organisms result in widespread morbidity and mortality throughout the world. Infectious disease is rampant in many locations, and unfortunately, numerous individuals suffer from more than one disease. Immunization of animals and humans with compositions comprising vaccines capable of generating an immune response is one of the only means of preventing or ameliorating the severity of disease.

Numerous diseases threaten not only humans but also animals. Influenza epidemics kill millions each year. Suppy of influenza vaccine is often inadequate for

the needs of numerous individuals. Further, the influenza vaccine is often designed on the best guesses of infectious disease experts concerning the expected antigens and strains for any given year. Different governments and the Centers for Disease Control face a major challenge in redesigning a vaccine and rapidly producing it to supply those in need.

Recent bioterrorism, for example the anthrax attacks, demonstrates the need for a stockpile of antidote. Bioterrorism against animals may also threaten the food supplies of nations. Engineered organisms provide a particular challenge for nations to rapidly adjust their vaccine production to attack newly discovered epitopes in these organisms.

Recent events have demonstrated the vulnerability of the U.S. to attack with biowarfare agents. For many of the agents listed as select agents by the Centers for Disease Control, vaccines are not readily available for humans or domestic livestock. Despite intense research with many of these organisms, the potential for a vaccine in the immediate future is not likely. In addition, treatment with antibiotics may or may not be effective, especially if a bacterium has been engineered to be multi-drug resistant.

Accordingly what is needed in a method for rapid and economical production of proteins, polypeptides, peptides and mixtures thereof that may be used as vaccines to prevent or ameliorate disease.

### SUMMARY OF THE INVENTION

5

10

15

20

25

30

The present invention provides a new, effective and efficient method of producing proteins, polypeptides and peptides (hereinafter generally "proteins") in a transgenic individual that may be used as vaccines. The present invention provides the capability to transform the genetic machinery of individuals in order to produce vaccines. Such transformation may be germ-line transformation and/or non germ-line transformation. The present invention facilitates production of one or more protein, polypeptide and peptide for use in vaccines.

This invention provides polynucleotide cassettes containing at least one gene of interest and one or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro nucleotide sequence. Each of the genes of interest encodes a protein that may be used in the vaccine. One discovery of the present

5

10

15

20

25

30

invention is the use of pro portions of prepro signal sequences to facilitate appropriate processing, expression, and/or formation of proteins in an individual. Preferred pro polynucleotides are provided in SEQ ID NO:1 and SEQ ID NO:2.

Another discovery of the present invention is that cecropin prepro sequences facilitate appropriate processing, expression, and/or formation of desired proteins in an individual. Accordingly, the present invention includes polynucleotide cassettes containing at least one gene of interest operably-linked to a cecropin prepro sequence. Preferred cecropin prepro polynucleotides are provided in SEQ ID NO:3 and SEQ ID NO:4. The present invention also includes polynucleotide cassettes containing at least one gene of interest operably linked to a cecropin prepro polynucleotide, wherein pro sequences are located between the at least one gene of interest when more than one gene of interest is present.

A non-limiting list of the diseases and organisms that may be targets of the vaccines of the present invention follows: Anthrax (Bacillus anthracis), Bacillus anthracis (anthrax), Botulism (Clostridium botulinum toxin), Brucella species (brucellosis), Brucellosis (Brucella species), Burkholderia mallei (glanders), Burkholderia pseudomallei (melioidosis), Chlamydia psittaci (psittacosis), Cholera (Vibrio cholerae), Clostridium botulinum toxin (botulism), Clostridium perfringens (Epsilon toxin), Coxiella burnetii (Q fever), E. coli O157:H7 (Escherichia coli), Emerging infectious diseases such as Nipah virus and hantavirus, Epsilon toxin of Clostridium perfringens, Escherichia coli O157:H7 (E. coli), Food safety threats (e.g., Salmonella species, Escherichia coli O157:H7, Shigella), Francisella tularensis (tularemia), Glanders (Burkholderia mallei), Melioidosis (Burkholderia pseudomallei), Plague (Yersinia pestis), Psittacosis (Chlamydia psittaci), Q fever (Coxiella burnetii), Ricin toxin from Ricinus communis (castor beans), Rickettsia prowazekii (typhus fever), Salmonella species (salmonellosis), Salmonella Typhi (typhoid fever), Salmonellosis (Salmonella species), Shigella (shigellosis), Shigellosis (Shigella), Smallpox (variola major), Staphylococcal enterotoxin B, Tularemia (Francisella tularensis), Typhoid fever (Salmonella Typhi), Typhus fever (Rickettsia prowazekii), Variola major (smallpox), Vibrio cholerae (cholera), Viral encephalitis (alphaviruses [e.g., Venezuelan equine encephalitis, eastern equine encephalitis, western equine encephalitis]), Viral hemorrhagic fevers (filoviruses [e.g., Ebola, Marburg] and arenaviruses [e.g., Lassa, Machupo]), Water safety threats (e.g., Vibrio

5

10

15

20

25

30

cholerae, Cryptosporidium parvum), and Yersinia pestis (plague). Other diseases and organisms that may be targets of the vaccines of the present invention are described in Harrison's Principles of Internal Medicine, (Fauci et al. eds, 14<sup>th</sup> edition, McGraw Hill, New York, USA (1998)).

These polynucleotide cassettes are administered to an individual for expression of polypeptide sequences and the formation of a desired protein or multimeric protein for use as a vaccine. Preferably, the individual is an animal from which the multimeric protein can be harvested. In one embodiment, preferred animals are egg-laying or milk-producing animals.

In one embodiment, the egg-laying transgenic animal is an avian. The method of the present invention may be used in avians including Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Passeriformes, Coraciformes, Ralliformes, Cuculiformes, Columbiformes, Galliformes, Anseriformes, and Herodiones. Preferably, the egg-laying transgenic animal is a poultry bird. More preferably, the bird is a chicken, turkey, duck, goose or quail. Another preferred bird is a ratite, such as, an emu, an ostrich, a rhea, or a cassowary. Other preferred birds are partridge, pheasant, kiwi, parrot, parakeet, macaw, falcon, eagle, hawk, pigeon, cockatoo, song birds, jay bird, blackbird, finch, warbler, canary, toucan, mynah, or sparrow.

In some embodiments, the polynucleotide cassettes are located within transposon-based vectors that allow for incorporation of the cassettes into the DNA of the individual. The transposon-based vectors of the present invention include a transposase, operably-linked to a first promoter, and a coding sequence for a protein or peptide of interest operably-linked to a second promoter, wherein the coding sequence for the protein or peptide of interest and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes the following characteristics: a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the nucleotide at the third base position of each codon is changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of

5

10

15

20

25

30

the transposase gene. In some embodiments, the effective polyA sequence is an avian optimized polyA sequence.

In one embodiment, the transposon-based vector comprises an avian optimized polyA sequence and does not comprise a modified Kozak sequence comprising ACCATG (SEQ ID NO:5). One example of such a transposon-based vector is the pTnMCS vector (SEQ ID NO:56). In another embodiment the transposon-based vector comprises a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. One example of such a transposon-based vector is the pTnMod vector (SEQ ID NO:57).

Accordingly, it is an object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines.

It is another object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines, wherein the proteins are antibodies.

It is another object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines, wherein the proteins are multimeric proteins.

It is another object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines, wherein the proteins are monoclonal antibodies.

Yet another object of the present invention is to vaccinate individuals, particularly animals and humans, with the vaccines of the present invention in order to prevent or ameliorate disease.

Yet another object of the present invention is to vaccinate individuals, particularly animals and humans, with a vaccine comprising a combination of the proteins or multimeric proteins of the present invention, together with antibodies of the present invention, in order to prevent or ameliorate disease.

Another object of the present invention is to provide proteins, multimeric proteins and antibodies that are useful in the preparation of a medicament that may be used as a vaccine.

An advantage of the present invention is that the proteins are produced more efficiently, in greater quantities and less expensively than taught by prior art methods.

These and other objects, features and advantages of the present invention will become apparent after a review of the following detailed description of the disclosed embodiments and claims.

## 10 BRIEF DESCRIPTION OF THE FIGURES

5

15

20

25

30

Figure 1 depicts schematically a polynucleotide cassette containing two genes of interest operably-linked to two pro polynucleotides, wherein the first pro polynucleotide is a part of a prepro polynucleotide. "Prom" indicates promoter.

Figure 2 depicts schematically a polynucleotide cassette containing polynucleotides encoding for a heavy chain and a light chain of an antibody. "Oval prom" indicates an ovalbumin promoter. The polynucleotide cassette contains pro and prepro sequences and is flanked by insertion sequences (IS) recognized by a transposase.

Figure 3 depicts schematically a polynucleotide cassette containing a cecropin prepro sequence operably-linked to two genes of interest. Between the genes of interest resides a cleavage site indicates by "CS."

Figure 4 depicts schematically a polynucleotide cassette containing two genes of interest, a promoter (prom), a signal sequence (SS) and a cleavage site (CS). The polynucleotide cassette is flanked by insertion sequences (IS) recognized by a transposase.

Figure 5 is a picture of a gel showing partially purified egg white derived from a transgenic avian run under reducing and non-reducing conditions.

# DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a new, effective and efficient method of producing proteins in an individual that may be used as vaccines. Such proteins include polypeptides, peptides and mixtures thereof and are generally called "proteins" throughout the present application. One or more proteins may be made

with the present invention. Antibodies may be made with the present invention and used as vaccines, alone or in conjunction with other proteins and multimeric proteins. Multimeric proteins may be produced with the persent invention and include at least two proteins, wherein the at least two proteins may be the same of different. A multimeric protein includes associated multimeric proteins (two or more associated polypeptides) and multivalent multimeric proteins (a single polypeptide encoded by more than one gene of interest). Expression and/or formation of the multimeric protein in the individual is achieved by administering a polynucleotide cassette containing the gene or genes of interest to the individual. The polynucleotide cassette may additionally contain one or more pro sequences, prepro sequences, cecropin prepro sequences, and/or cleavage site sequences.

5

10

15

20

25

30

One or more genes of interest for incorporation into the polynucleotide cassette are chosen based on the genes that generally code for epitopes of organisms that are believed to be antigenic. Such coding sequences are generally known to one of ordinary skill in the art and can be obtained through GenBank and the published scientific and patent literature. It is within the scope of the present invention to search the genome of organisms to identify coding regios for particular proteins, polypeptides, peptides or epitopes that could serve as epitopes for vaccines. A protein may be made for use in a vaccine against any desired disease or disease producing organism using the method of the present invention. A non-limiting list of the diseases and organisms that may be targets of the vaccines of the present invention follows: Anthrax (Bacillus anthracis), Bacillus anthracis (anthrax), Botulism (Clostridium botulinum toxin), Brucella species (brucellosis), Brucellosis (Brucella species), Burkholderia mallei (glanders), Burkholderia pseudomallei (melioidosis), Chlamydia psittaci (psittacosis), Cholera (Vibrio cholerae), Clostridium botulinum toxin (botulism), Clostridium perfringens (Epsilon toxin), Coxiella burnetii (Q fever), E. coli O157:H7 (Escherichia coli), Emerging infectious diseases such as Nipah virus and hantavirus, Epsilon toxin of Clostridium perfringens, Escherichia coli O157:H7 (E. coli), Food safety threats (e.g., Salmonella species, Escherichia coli O157:H7, Shigella), Francisella tularensis (tularemia), Glanders (Burkholderia mallei), Melioidosis (Burkholderia pseudomallei), Plague (Yersinia pestis), Psittacosis (Chlamydia psittaci), Q fever (Coxiella burnetii), Ricin toxin from Ricinus communis (castor beans), Rickettsia prowazekii (typhus fever), Salmonella

(salmonellosis), Salmonella Typhi (typhoid fever), Salmonellosis (Salmonella species), Shigella (shigellosis), Shigellosis (Shigella), Smallpox (variola major), Staphylococcal enterotoxin B, Tularemia (Francisella tularensis), Typhoid fever (Salmonella Typhi), Typhus fever (Rickettsia prowazekii), Variola major (smallpox), Vibrio cholerae (cholera), Viral encephalitis (alphaviruses [e.g., Venezuelan equine encephalitis, eastern equine encephalitis, western equine encephalitis]), Viral hemorrhagic fevers (filoviruses [e.g., Ebola, Marburg] and arenaviruses [e.g., Lassa, Machupo]), Water safety threats (e.g., Vibrio cholerae, Cryptosporidium parvum), and Yersinia pestis (plague). Other diseases and organisms that may be targets of the vaccines of the present invention are described in Harrison's Principles of Internal Medicine, (Fauci et al. eds, 14<sup>th</sup> edition, McGraw Hill, New York, USA (1998)).

5

10

15

20

25

30

This invention provides polynucleotide cassettes containing at least one gene of interest and one or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro nucleotide sequence. When one gene of interest is present in the polynucleotide cassette, a single protein is made, as opposed to a The present invention provides multimeric protein as described below. polynucleotide cassettes containing two or more genes of interest and two or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro These polynucleotide cassettes are useful for making nucleotide sequence. multimeric proteins. Each of the genes of interest encodes a polypeptide that forms a part of the multimeric protein. These polynucleotide cassettes are administered to an individual for expression of the polypeptide sequences and expression and/or formation of the multimeric protein. Preferably, the individual is an animal from which the proteins or multimeric protein can be harvested. Preferred animals are egglaying or milk-producing animals. In some embodiments, the polynucleotide cassettes are located within transposon-based vectors that allow for incorporation of the cassettes into the DNA of the individual.

The pro polynucleotide sequences operably-linked to the genes of interest include pro portions of prepro polynucleotide sequences commonly associated with polynucleotides encoding proteins secreted from a cell in nature. It may be that the pre polynucleotide sequence functions to direct the resultant protein into the endoplasmic reticulum and the pro sequence is cleaved within the endoplasmic reticulum or Golgi complex of a cell containing the protein. While prepro

5

10

15

20

25

30

polynucleotide sequences are associated with secreted polypeptides in nature, one discovery of the present invention is the use of pro portions of the prepro signal sequences to facilitate appropriate processing, expression, and/or formation of proteins and multimeric proteins, and more particularly, associated multimeric proteins. In the present invention, each gene of interest is operably-linked with a pro polynucleotide sequence. Figure 1 shows schematically one polynucleotide cassette containing two genes of interest, wherein each gene of interest is operably-linked to a pro polynucleotide sequence. The first gene of interest is operably-linked to a pro polynucleotide sequence that is part of a prepro polynucleotide sequence, while the second gene of interest is operably-linked to a pro polynucleotide sequence that is not part of a prepro polynucleotide sequence, but may have been derived from a prepro polynucleotide sequence. Accordingly, the term "pro sequence" encompasses a pro sequence that is part of a prepro sequence and a pro sequence that is not part of a prepro sequence, but may have been derived from a prepro sequence. In preferred embodiments, the most 5' pro polynucleotide sequence in the polynucleotide cassette is a part of a prepro polynucleotide sequence.

Several examples of prepro polynucleotides from which a pro polynucleotide can be derived or be a part of are a cecropin prepro, lysozyme prepro, ovomucin prepro, ovotransferrin prepro, a signal peptide for tumor necrosis factor receptor (SEQ ID NO:6), a signal peptide encoded by a polynucleotide sequence provided in one of SEQ ID NO:7-54 and a signal peptide provide in SEQ ID NO:55. The prepro or pro polynucleotide can be a cecropin prepro or pro polynucleotide selected from the group consisting of cecropin A1, cecropin A2, cecropin B, cecropin C, cecropin D, cecropin E and cecropin F. In a preferred embodiment, the pro polynucleotide is a cecropin B pro polynucleotide having a sequence shown in SEQ ID NO:1 or SEQ ID NO:2. A preferred prepro polynucleotide is a cecropin B polynucleotide having a sequence shown in SEQ ID NO:3 or SEQ ID NO:4.

Figure 1 provides one embodiment of the invention wherein the polynucleotide cassette includes two genes of interest and two pro polynucleotide sequences arranged in the following order: a prepro polynucleotide, a first gene of interest, a pro polynucleotide, and a second gene of interest. Preferably, the sequences are arranged in the aforementioned order beginning at a 5' end of the polynucleotide cassette. Figure 2 provides a more specific embodiment of the present

invention wherein the first and second genes of interest are polynucleotides encoding antibody heavy and light chains. However, the invention includes polynucleotide cassettes containing at least two genes of interest. Each of the genes of interest is operably-linked to a pro polynucleotide. Each of these pro polynucleotides can be the same, or each can be different. In one embodiment, all of the pro polynucleotides in the polynucleotide cassette are the same and are cecropin pro polynucleotides. The most 5' cecropin pro polynucleotide is preferably a part of a cecropin prepro polynucleotide sequence as shown in Figure 3.

5

10

15

20

25

30

The polynucleotide cassettes of the present invention may be administered to an individual for production of a protein or a multimeric protein in that individual. The discussion that follows is intented to encompass methods for making one protein or a multimeric protein. Accordingly, the present invention includes a method of producing a protein in an individual comprising administering to the individual a polynucleotide cassette comprising one gene of interest encoding the protein, wherein the gene of interest is operably-linked to a pro polynucleotide sequence. The present invention includes a method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide cassette comprising at least two genes of interest, each encoding a part of the multimeric protein, wherein each gene of interest is operably-linked to a pro polynucleotide sequence. The present invention also includes a method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide cassette comprising a cecropin prepro sequence operably-linked to two or more genes of interest, each gene of interest encoding a part of the multimeric protein. This second method does not require the linking of pro polynucleotides to each gene of interest since the use of a cecropin prepro sequence itself in a polynucleotide cassette facilitates processing, Polynucleotide cassettes expression, and/or formation of multimeric proteins. containing the cecropin prepro polynucleotide can contain one or at least two genes of interest. Preferably, the cecropin prepro polynucleotide is located 5' of the one or more genes of interest in the polynucleotide cassette. One exemplary polynucleotide cassette is shown in Figure 3. In a preferred embodiment, the prepro sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4. As shown in Figure 3, the polynucleotide cassettes containing a cecropin prepro polynucleotide preferably contain a cleavage site between each of two genes of interest. Such cleavage site(s)

5

10

15

20

25

30

may be nucleotides encoding any cleavage sites including, but not limited to, an enzymatic cleavage site, a pro polynucleotide, and a photolabile cleavage site, a chemical cleavage site, and a self-splicing cleavage site (i.e., intein). Cleavage sites are discussed in more detail below.

The polynucleotide cassettes of the present invention are particularly suited for production of proteins and/or multimeric proteins in an individual. Individuals include both humans and animals. Preferred animals are egg-laying animals and milk-producing animals. As used herein, the term "egg-laying animal" includes all amniotes such as birds, turtles, lizards and monotremes. Monotremes are egg-laying mammals and include the platypus and echidna. The term "bird" or "fowl," as used herein, is defined as a member of the Aves class of animals which are characterized as warm-blooded, egg-laying vertebrates primarily adapted for flying. Avians include, without limitation, Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Passeriformes. Coraciformes, Ralliformes, Cuculiformes, Columbiformes, Galliformes, Anseriformes, and Herodiones. The term "Ratite," as used herein, is defined as a group of flightless, mostly large, running birds comprising several orders and including the emus, ostriches, kiwis, and cassowaries. The term "Psittaciformes", as used herein, includes parrots and refers to a monofamilial order of birds that exhibit zygodactylism and have a strong hooked bill. A "parrot" is defined as any member of the avian family Psittacidae (the single family of the Psittaciformes), distinguished by the short, stout, strongly hooked beak. Preferred avians are poultry birds including chickens, quail, turkeys, geese and ducks. The term "chicken" as used herein denotes chickens used for table egg production, such as egg-type chickens, chickens reared for public meat consumption, or broilers, and chickens reared for both egg and meat production ("dual-purpose" chickens). The term "chicken" also denotes chickens produced by primary breeder companies, or chickens that are the parents, grandparents, great-grandparents, etc. of those chickens reared for public table egg, meat, or table egg and meat consumption.

When the polynucleotide cassettes of the present invention are administered to an egg-laying or milk-producing animal, a transgenic animal containing a polynucleotide cassette is created and the animal produces a transgenic protein or multimeric protein. It is preferred that the resultant protein or multimeric protein is deposited in the egg or in the milk. Various different signal sequences and promoters

may be used to achieve deposition of the protein or multimeric protein in the egg or in the milk and these are described in more detail below. In order to achieve a transgenic animal containing a polynucleotide cassette of the present invention, the polynucleotide cassettes can be administered to the individual with, or contained in, any vector, as naked DNA, or in any delivery construct or solution commonly known to one of ordinary skill in the art. A preferred vector for incorporation of the polynucleotide cassettes into an individual is a transposon-based vector described below.

#### **Definitions**

5

10

15

20

25

30

It is to be understood that as used in the specification and in the claims, "a" or "an" can mean one or more, depending upon the context in which it is used. Thus, for example, reference to "a cell" can mean that at least one cell can be utilized.

The term protein includes polypeptide and peptide and these terms are generally encompassed under the term "protein" and are also used interchangeablt herein. A multimeric protein includes more than one protein, polypeptide or peptide and is defined further below.

The term "antibody" is used interchangeably with the term "immunoglobulin" and is defined herein as a protein synthesized by an animal or a cell of the immune system in response to the presence of a foreign substance commonly referred to as an "antigen" or an "immunoger.". The term antibody includes fragments of antibodies. Antibodies are characterized by specific affinity to a site on the antigen, wherein the site is referred to an "antigenic determinant" or an "epitope". Antigens can be naturally occurring or artificially engineered. Preferred antigens are those from organisms against which a vaccine is desired. Antigens and suspected or known epitopes may be obtained from any organism, including but not limited to those recited herein.

Artificially engineered antigens include but are not limited to small molecules, such as small peptides, attached to haptens such as macromolecules, for example proteins, nucleic acids, or polysaccharides. Artificially designed or engineered variants of naturally occurring antibodies and artificially designed or engineered antibodies not occurring in nature are all included in the current definition. Such variants include conservatively substituted amino acids and other forms of substitution as described in the section concerning proteins and polypeptides.

The term "egg" is defined herein as including a large female sex cell enclosed in a porous, calcarous or leathery shell, produced by birds and reptiles. The term "ovum" is defined as a female gamete, and is also known as an egg. Therefore, egg production in all animals other than birds and reptiles, as used herein, is defined as the production and discharge of an ovum from an ovary, or "ovulation". Accordingly, it is to be understood that the term "egg" as used herein is defined as a large female sex cell enclosed in a porous, calcarous or leathery shell, when a bird or reptile produces it, or it is an ovum when it is produced by all other animals.

5

10

15

20

25

30

The term "gene" is defined herein to include a polynucleotide that includes a coding region for a protein, peptide or polypeptide, with or without intervening sequences such as introns.

The term "multimeric protein" is defined herein to include one or more polypeptides that are associated, or joined, by any means including disulfde bonds. An example of this type of multimeric protein is an antibody that contains both heavy and light chains that are associated by disulfide bonds. These multimeric proteins are referred to herein as "associated multimeric proteins." The term "multimeric protein" also includes a polypeptide that is encoded by more than one gene of interest, such multimeric protein containg desired amino acid sequences encoded by the more than one gene of interest. An example of this type of multimeric protein is a single polypeptide containing a heavy chain polypeptide (first polypeptide of interest) and a light chain polypeptide (second polypeptide of interest). In these embodiments, the different polypeptides of interest may be separated by other polypeptide sequences such as spacer polypeptides and cleavage site polypeptides. These types of multimeric proteins are referred to herein as "multivalent multimeric proteins."

The term "milk-producing animal" refers herein to mammals including, but not limited to, bovine, ovine, porcine, equine, and primate animals. Milk-producing animals include but are not limited to cows, llamas, camels, goats, reindeer, zebu, water buffalo, yak, horses, pigs, rabbits, non-human primates, and humans.

The term "transgenic animal" refers to an animal having at least a portion of the transposon-based vector DNA is incorporated into its DNA. While a transgenic animal includes an animal wherein the transposon-based vector DNA is incorporated into the germline DNA, a transgenic animal also includes an animal having DNA in one or more somatic cells that contain a portion of the transposon-based vector DNA

for any period of time. In a preferred embodiment, a portion of the transposon-based vector comprises a gene of interest. More preferably, the gene of interest is incorporated into the animal's DNA for a period of at least five days, more preferably the laying life of the animal, and most preferably the life of the animal. In a further preferred embodiment, the animal is an avian.

The term "vector" is used interchangeably with the terms "construct", "DNA construct" and "genetic construct" to denote synthetic nucleotide sequences used for manipulation of genetic material, including but not limited to cloning, subcloning, sequencing, or introduction of exogenous genetic material into cells, tissues or organisms, such as birds. It is understood by one skilled in the art that vectors may contain synthetic DNA sequences, naturally occurring DNA sequences, or both. The vectors of the present invention are transposon-based vectors as described herein.

When referring to two nucleotide sequences, one being a regulatory sequence, the term "operably-linked" is defined herein to mean that the two sequences are associated in a manner that allows the regulatory sequence to affect expression of the other nucleotide sequence. It is not required that the operably-linked sequences be directly adjacent to one another with no intervening sequence(s).

The term "regulatory sequence" is defined herein as including promoters, enhancers and other expression control elements such as polyadenylation sequences, matrix attachment sites, insulator regions for expression of multiple genes on a single construct, ribosome entry/attachment sites, introns that are able to enhance expression, and silencers.

## Transposon-Based Vectors

5

10

15

20

25

30

While not wanting to be bound by the following statement, it is believed that the nature of the DNA construct is an important factor in successfully producing transgenic animals. The "standard" types of plasmid and viral vectors that have previously been almost universally used for transgenic work in all species, especially avians, have low efficiencies and may constitute a major reason for the low rates of transformation previously observed. The DNA (or RNA) constructs previously used often do not integrate into the host DNA, or integrate only at low frequencies. Other factors may have also played a part, such as poor entry of the vector into target cells. The present invention provides transposon-based vectors that can be administered to an animal that overcome the prior art problems relating to low transgene integration

5

10

15

20

25

30

frequencies. Two preferred transposon-based vectors of the present invention in which a transposase, gene of interest and other polynucleotide sequences may be introduced are termed pTnMCS (SEQ ID NO:56) and pTnMod (SEQ ID NO:57).

The transposon-based vectors of the present invention produce integration frequencies an order of magnitude greater than has been achieved with previous vectors. More specifically, intratesticular injections performed with a prior art transposon-based vector (described in U.S. Patent No. 5,719,055) resulted in 41% sperm positive roosters whereas intratesticular injections performed with the novel transposon-based vectors of the present invention resulted in 77% sperm positive roosters. Actual frequencies of integration were estimated by either or both comparative strength of the PCR signal from the sperm and histological evaluation of the testes and sperm by quantitative PCR.

The transposon-based vectors of the present invention include a transposase gene operably-linked to a first promoter, and a coding sequence for a desired protein or peptide operably-linked to a second promoter, wherein the coding sequence for the desired protein or peptide and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes one or more of the following characteristics: a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of one or more of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. In one embodiment, the transposon-based vector comprises an avian optimized polyA sequence and does not comprise a modified Kozak sequence comprising ACCATG (SEQ ID NO:5). One example of such a transposon-based vector is the pTnMCS vector (SEQ ID NO:56). In another embodiment the transposon-based vector comprises a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without

changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. One example of such a transposon-based vector is the pTnMod vector (SEQ ID NO:57). The transposon-based vector may additionally or alternatively include one or more of the following Kozak sequences at the 3' end of any promoter, including the promoter operably-linked to the transposase: ACCATGG (SEQ ID NO:58), AAGATGT (SEQ ID NO:59), ACGATGA (SEQ ID NO:60), AAGATGG (SEQ ID NO:61), GACATGA (SEQ ID NO:62), ACCATGA (SEQ ID NO:63), and ACCATGA (SEQ ID NO:64), ACCATGT (SEQ ID NO:65).

### Transposases and Insertion Sequences

5

10

15

20

25

30

In a further embodiment of the present invention, the transposase found in the transposase-based vector is an altered target site (ATS) transposase and the insertion sequences are those recognized by the ATS transposase. However, the transposase located in the transposase-based vectors is not limited to a modified ATS transposase and can be derived from any transposase. Transposases known in the prior art include those found in AC7, Tn5SEQ1, Tn916, Tn951, Tn1721, Tn 2410, Tn1681, Tn1, Tn2, Tn3, Tn4, Tn5, Tn6, Tn9, Tn10, Tn30, Tn101, Tn903, Tn501, Tn1000 (γδ), Tn1681, Tn2901, AC transposons, Mp transposons, Spm transposons, En transposons, Dotted transposons, Mu transposons, Ds transposons, dSpm transposons and I transposons. According to the present invention, these transposases and their regulatory sequences are modified for improved functioning as follows: a) the addition one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the promoter operably-linked to the transposase; b) a change of one or more of the codons for the first several amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) the addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) the addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene.

Although not wanting to be bound by the following statement, it is believed that the modifications of the first several N-terminal codons of the transposase gene facilitate transcription of the transposase gene, in part, by increasing strand dissociation during transcription. It is preferable that one or more of between

approximately the first 1 to 20, more preferably 3 to 15, and most preferably between 4 to 12 N-terminal codons of the transposase are modified such that the third base of each codon is changed to an A or a T without changing the encoded amino acid. In one embodiment, the first ten N-terminal codons of the transposase gene are modified in this manner. It is also preferred that the transposase contain mutations that make it less specific for preferred insertion sites and thus increases the rate of transgene insertion as discussed in U.S. Patent No. 5,719,055.

5

10

15

20

25

30

In some embodiments, the transposon-based vectors are optimized for expression in a particular host by changing the methylation patterns of the vector DNA. For example, prokaryotic methylation may be reduced by using a methylation deficient organism for production of the transposon-based vector. The transposon-based vectors may also be methylated to resemble eukaryotic DNA for expression in a eukaryotic host.

Transposases and insertion sequences from other analogous eukaryotic transposon-based vectors that can also be modified and used are, for example, the Drosophila P element derived vectors disclosed in U.S. Patent No. 6,291,243; the Drosophila mariner element described in Sherman et al. (1998); or the sleeping beauty transposon. See also Hackett et al. (1999); D. Lampe et al., 1999. Proc. Natl. Acad. Sci. USA, 96:11428-11433; S. Fischer et al., 2001. Proc. Natl. Acad. Sci. USA, 98:6759-6764; L. Zagoraiou et al., 2001. Proc. Natl. Acad. Sci. USA, 98:11474-11478; and D. Berg et al. (Eds.), Mobile DNA, Amer. Soc. Microbiol. (Washington, D.C., 1989). However, it should be noted that bacterial transposon-based elements are preferred, as there is less likelihood that a eukaryotic transposase in the recipient species will recognize prokaryotic insertion sequences bracketing the transgene.

Many transposases recognize different insertion sequences, and therefore, it is to be understood that a transposase-based vector will contain insertion sequences recognized by the particular transposase also found in the transposase-based vector. In a preferred embodiment of the invention, the insertion sequences have been shortened to about 70 base pairs in length as compared to those found in wild-type transposons that typically contain insertion sequences of well over 100 base pairs.

While the examples provided below incorporate a "cut and insert" Tn10 based vector that is destroyed following the insertion event, the present invention also encompasses the use of a "rolling replication" type transposon-based vector. Use of a

PCT/US2003/041269 WO 2004/067743

rolling replication type transposon allows multiple copies of the transposon/transgene to be made from a single transgene construct and the copies inserted. This type of transposon-based system thereby provides for insertion of multiple copies of a transgene into a single genome. A rolling replication type transposon-based vector may be preferred when the promoter operably-linked to gene of interest is endogenous to the host cell and present in a high copy number or highly expressed. However, use of a rolling replication system may require tight control to limit the insertion events to non-lethal levels. Tn1, Tn2, Tn3, Tn4, Tn5, Tn9, Tn21, Tn501, Tn551, Tn951, Tn1721, Tn2410 and Tn2603 are examples of a rolling replication type transposon, although Tn5 could be both a rolling replication and a cut and insert type transposon.

Stop Codons and PolyA Sequences

5

10

15

20

25

30

In one embodiment, the transposon-based vector contains two stop codons operably-linked to the transposase and/or to the gene of interest. In an alternate embodiment, one stop codon of UAA or UGA is operably linked to the transposase and/or to the gene of interest.

As used herein an "effective polyA sequence" refers to either a synthetic or non-synthetic sequence that contains multiple and sequential nucleotides containing an adenine base (an A polynucleotide string) and that increases expression of the gene to which it is operably-linked. A polyA sequence may be operably-linked to any gene in the transposon-based vector including, but not limited to, a transposase gene and a gene of interest. A preferred polyA sequence is optimized for use in the host animal or human. In one embodiment, the polyA sequence is optimized for use in an avian species and more specifically, a chicken. An avian optimized polyA sequence generally contains a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 75 base pairs that precede the A polynucleotide string and thereby separate the stop codon from the A In one embodiment of the present invention, the polyA polynucleotide string. sequence comprises a conalbumin polyA sequence as provided in SEQ ID NO:66 and as taken from GenBank accession # Y00407, base pairs 10651-11058. In another embodiment, the polyA sequence comprises a synthetic polynucleotide sequence shown in SEQ ID NO:67. In yet another embodiment, the polyA sequence comprises an avian optimized polyA sequence provided in SEQ ID NO:68. A chicken optimized

polyA sequence may also have a reduced amount of CT repeats as compared to a synthetic polyA sequence.

It is a suprising discovery of the present invention that such an avian optimized poly A sequence increases expression of a polynucleotide to which it is operably-linked in an avian as compared to a non-avian optimized polyA sequence. Accordingly, the present invention includes methods of or increasing incorporation of a gene of interest wherein the gene of interest resides in a transposon-based vector containing a transposase gene and wherein the transposase gene is operably linked to an avian optimized polyA sequence. The present invention also includes methods of increasing expression of a gene of interest in an avian that includes administering a gene of interest to the avian, wherein the gene of interest is operably-linked to an avian optimized polyA sequence. An avian optimized polyA nucleotide string is defined herein as a polynucleotide containing an A polynucleotide string and a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 75 base pairs that precede the A polynucleotide string. The present invention further provides transposon-based vectors containing a gene of interest or transposase gene operably linked to an avian optimized polyA sequence.

### Promoters and Enhancers

10

15

20

25

30

The first promoter cerably-linked to the transposase gene and the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. Constitutive promoters include, but are not limited to, immediate early cytomegalovirus (CMV) promoter, herpes simplex virus 1 (HSV1) immediate early promoter, SV40 promoter, lysozyme promoter, early and late CMV promoters, early and late HSV promoters, β-actin promoter, tubulin promoter, Rous-Sarcoma virus (RSV) promoter, and heat-shock protein (HSP) promoter. Inducible promoters include tissue-specific promoters, developmentally-regulated promoters and chemically inducible promoters. Examples of tissue-specific promoters include the glucose 6 phosphate (G6P) promoter, vitellogenin promoter, ovalbumin promoter, ovomucoid promoter, conalbumin promoter, ovotransferrin promoter, prolactin promoter, kidney uromodulin promoter, and placental lactogen promoter. In one embodiment, the vitellogenin promoter includes a polynucleotide sequence of SEQ ID NO:69. The G6P promoter sequence may be deduced from a rat G6P gene

5

10

15

20

25

30

untranslated upstream region provided in GenBank accession number U57552.1. Examples of developmentally-regulated promoters include the homeobox promoters and several hormone induced promoters. Examples of chemically inducible promoters include reproductive hormone induced promoters and antibiotic inducible promoters such as the tetracycline inducible promoter and the zinc-inducible metallothionine promoter.

Other inducible promoter systems include the Lac operator repressor system inducible by IPTG (isopropyl beta-D-thiogalactoside) (Cronin, A. et al. 2001. Genes and Development, v. 15), ecdysone-based inducible systems (Hoppe, U. C. et al. 2000. Mol. Ther. 1:159-164); estrogen-based inducible systems (Braselmann, S. et al. 1993. Proc. Natl. Acad. Sci. 90:1657-1661); progesterone-based inducible systems using a chimeric regulator, GLVP, which is a hybrid protein consisting of the GAL4 binding domain and the herpes simplex virus transcriptional activation domain, VP16, and a truncated form of the human progesterone receptor that retains the ability to bind ligand and can be turned on by RU486 (Wang, et al. 1994. Proc. Natl. Acad. Sci. 91:8180-8184); CID-based inducible systems using chemical inducers of dimerization (CIDs) to regulate gene expression, such as a system wherein rapamycin induces dimerization of the cellular proteins FKBP12 and FRAP (Belshaw, P. J. et al. 1996. J. Chem. Biol. 3:731-738; Fan, L. et al. 1999. Hum. Gene Ther. 10:2273-2285; Shariat, S.F. et al. 2001. Cancer Res. 61:2562-2571; Spencer, D.M. 1996. Curr. Biol. 6:839-847). Chemical substances that activate the chemically inducible promoters can be administered to the animal containing the transgene of interest via any method known to those of skill in the art.

Other examples of cell or tissue-specific and constitutive promoters include but are not limited to smooth-muscle SM22 promoter, including chimeric SM22alpha/telokin promoters (Hoggatt A.M. et al., 2002. Circ Res. 91(12):1151-9); ubiquitin C promoter (Biochim Biophys Acta, 2003. Jan. 3;1625(1):52-63); Hsf2 promoter; murine COMP (cartilage oligomeric matrix protein) promoter; early B cell-specific mb-1 promoter (Sigvardsson M., et al., 2002. Mol. Cell Biol. 22(24):8539-51); prostate specific antigen (PSA) promoter (Yoshimura I. et al., 2002, J. Urol. 168(6):2659-64); exorh promoter and pineal expression-promoting element (Asaoka Y., et al., 2002. Proc. Natl. Acad. Sci. 99(24):15456-61); neural and liver ceramidase gene promoters (Okino N. et al., 2002. Biochem. Biophys. Res. Commun.

299(1):160-6); PSP94 gene promoter/enhancer (Gabril M.Y. et al., 2002. Gene Ther. 9(23):1589-99); promoter of the human FAT/CD36 gene (Kuriki C., et al., 2002. Biol. Pharm. Bull. 25(11):1476-8); VL30 promoter (Staplin W.R. et al., 2002. Blood October 24, 2002); and IL-10 promoter (Brenner S., et al., 2002. J. Biol. Chem. December 18, 2002).

5

10

15

20

25

30

Examples of avian promoters include, but are not limited to, promoters controlling expression of egg white proteins, such as ovalbumin, ovotransferrin (conalbumin), ovomucoid, lysozyme, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, ovostatin (ovomacroglobin), cystatin, avidin, thiamine-binding protein, glutamyl aminopeptidase minor glycoprotein 1, minor glycoprotein 2; and promoters controlling expression of egg-yolk proteins, such as vitellogenin, very lowdensity lipoproteins, low density lipoprotein, cobalamin-binding protein, riboflavinbinding protein, biotin-binding protein (Awade, 1996. Z. Lebensm. Unters. Forsch. 202:1-14). An advantage of using the vitellogenin promoter is that it is active during the egg-laying stage of an animal's life-cycle, which allows for the production of the protein of interest to be temporally connected to the import of the protein of interest into the egg yolk when the protein of interest is equipped with an appropriate targeting sequence. In some embodiments, the avian promoter is an oviduct-specific promoter. As used herein, the term "oviduct-specific promoter" includes, but is not limited to, ovalbumin, ovotransferrin (conalbumin), ovomucoid, lysozyme, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, and ovostatin (ovomacroglobin) promoters.

Liver-specific promoters of the present invention include, but are not limited to, the following promoters, vitellogenin promoter, G6P promoter, cholesterol-7-alpha-hydroxylase (CYP7A) promoter, phenylalanine hydroxylase (PAH) promoter, protein C gene promoter, insulin-like growth factor I (IGF-I) promoter, bilirubin UDP-glucuronosyltransferase promoter, aldolase B promoter, furin promoter, metallothioneine promoter, albumin promoter, and insulin promoter.

Also included in the present invention are promoters that can be used to target expression of a protein of interest into the milk of a milk-producing animal including, but not limited to,  $\beta$  lactoglobin promoter, whey acidic protein promoter, lactalbumin promoter and casein promoter.

5

10

15

20

25

30

Promoters associated with cells of the immune system may also be used. Acute phase promoters such as interleukin (IL)-1 and IL-2 may be employed. Promoters for heavy and light chain Ig may also be employed. The promoters of the T cell receptor components CD4 and CD8, B cell promoters and the promoters of CR2 (complement receptor type 2) may also be employed. Immune system promoters are preferably used when the desired protein is an antibody protein.

Also included in this invention are modified promoters/enhancers wherein elements of a single promoter are duplicated, modified, or otherwise changed. In one embodiment, steroid hormone-binding domains of the ovalbumin promoter are moved from about -6.5 kb to within approximately the first 1000 base pairs of the gene of interest. Modifying an existing promoter with promoter/enhancer elements not found naturally in the promoter, as well as building an entirely synthetic promoter, or drawing promoter/enhancer elements from various genes together on a non-natural backbone, are all encompassed by the current invention.

Accordingly, it is to be understood that the promoters contained within the transposon-based vectors of the present invention may be entire promoter sequences or fragments of promoter sequences. For example, in one embodiment, the promoter operably linked to a gene of interest is an approximately 900 base pair fragment of a chicken ovalbumin promoter (SEQ ID NO:70). The constitutive and inducible promoters contained within the transposon-based vectors may also be modified by the addition of one or more modified Kozak sequences of ACCATG (SEQ ID NO:5).

As indicated above, the present invention includes transposon-based vectors containing one or more enhancers. These enhancers may or may not be operably-linked to their native promoter and may be located at any distance from their operably-linked promoter. A promoter operably-linked to an enhancer is referred to herein as an "enhanced promoter." The enhancers contained within the transposon-based vectors are preferably enhancers found in birds, and more preferably, an ovalbumin enhancer, but are not limited to these types of enhancers. In one embodiment, an approximately 675 base pair enhancer element of an ovalbumin promoter is cloned upstream of an ovalbumin promoter with 300 base pairs of spacer DNA separating the enhancer and promoter. In one embodiment, the enhancer used as a part of the present invention comprises base pairs 1-675 of a Chicken Ovalbumin

enhancer from GenBank accession #S82527.1. The polynucleotide sequence of this enhancer is provided in SEQ ID NO:71.

Also included in some of the transposon-based vectors of the present invention are cap sites and fragments of cap sites. In one embodiment, approximately 50 base pairs of a 5' untranslated region wherein the capsite resides are added on the 3' end of an enhanced promoter or promoter. An exemplary 5' untranslated region is provided in SEQ ID NO:72. A putative cap-site residing in this 5' untranslated region preferably comprises the polynucleotide sequence provided in SEQ ID NO:73.

5

10

15

20

25

30

In one embodiment of the present invention, the first promoter operably-linked to the transposase gene is a constitutive promoter and the second promoter operably-linked to the gene of interest is a tissue-specific promoter. In the second embodiment, use of the first constitutive promoter allows for constitutive activation of the transposase gene and incorporation of the gene of interest into virtually all cell types, including the germline of the recipient animal. Although the gene of interest is incorporated into the germline generally, the gene of interest is only expressed in a tissue-specific manner. A transposon-based vector having a constitutive promoter operably-linked to the transposase gene can be administered by any route, and in one embodiment, the vector is administered to an ovary or to an artery leading to the ovary. In another embodiment, the vector is administered into the lumen of the oviduct or into an artery supplying the oviduct.

It should be noted that cell- or tissue-specific expression as described herein does not require a complete absence of expression in cells or tissues other than the preferred cell or tissue. Instead, "cell-specific" or "tissue-specific" expression refers to a majority of the expression of a particular gene of interest in the preferred cell or tissue, respectively.

When incorporation of the gene of interest into the germline is not preferred, the first promoter operably-linked to the transposase gene can be a tissue-specific promoter. For example, transfection of a transposon-based vector containing a transposase gene operably-linked to an oviduct specific promoter such as the ovalbumin promoter provides for activation of the transposase gene and incorporation of the gene of interest in the cells of the oviduct but not into the germline and other cells generally. In this embodiment, the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. In a preferred

embodiment, both the first promoter and the second promoter are an ovalbumin promoter. In embodiments wherein tissue-specific expression or incorporation is desired, it is preferred that the transposon-based vector is administered directly to the tissue of interest or to an artery leading to the tissue of interest. In a preferred embodiment, the tissue of interest is the oviduct and administration is achieved by direct injection into the lumen of the oviduct or an artery leading to the oviduct. In a further preferred embodiment, administration is achieved by direct injection into the lumen of the magnum or the infundibulum of the oviduct.

Accordingly, cell specific promoters may be used to enhance transcription in selected tissues. In birds, for example, promoters that are found in cells of the fallopian tube, such as ovalbumin, conalbumin, ovomucoid and/or lysozyme, are used in the vectors to ensure transcription of the gene of interest in the epithelial cells and tubular gland cells of the fallopian tube, leading to synthesis of the desired protein encoded by the gene and deposition into the egg white. In mammals, promoters specific for the epithelial cells of the alveoli of the mammary gland, such as prolactin, insulin, beta lactoglobin, whey acidic protein, lactalbumin, casein, and/or placental lactogen, are used in the design of vectors used for transfection of these cells for the production of desired proteins for deposition into the milk. In liver cells, the G6P promoter may be employed to drive transcription of the gene of interest for protein production. Proteins made in the liver of birds may be delivered to the egg yolk.

In order to achieve higher or more efficient expression of the transposase gene, the promoter and other regulatory sequences operably-linked to the transposase gene may be those derived from the host. These host specific regulatory sequences can be tissue specific as described above or can be of a constitutive nature. For example, an avian actin promoter and its associated polyA sequence can be operably-linked to a transposase in a transposase-based vector for transfection into an avian. Examples of other host specific promoters that could be operably-linked to the transposase include the myosin and DNA or RNA polymerase promoters.

### Directing Sequences

5

10

15

20

25

30

In some embodiments of the present invention, the gene of interest is operably-linked to a directing sequence or a sequence that provides proper conformation to the desired protein encoded by the gene of interest. As used herein, the term "directing sequence" refers to both signal sequences and targeting sequences.

An egg directing sequence includes, but is not limited to, an ovomucoid signal sequence, an ovalbumin signal sequence, a cecropin pre pro signal sequence, and a vitellogenin targeting sequence. The term "signal sequence" refers to an amino acid sequence, or the polynucleotide sequence that encodes the amino acid sequence, a portion or the entirety of which directs the protein to which it is linked to the endoplasmic reticulum in a eukaryote, and more preferably the translocational pores in the endoplasmic reticulum, or the plasma membrane in a prokaryote, or mitochondria, such as for the purpose of gene therapy for mitochondrial diseases. Signal and targeting sequences can be used to direct a desired protein into, for example, the milk, when the transposon-based vectors are administered to a milk-producing animal.

Signal sequences can also be used to direct a desired protein into, for example, a secretory pathway for incorporation into the egg yolk or the egg white, when the transposon-based vectors are administered to a bird or other egg-laying animal. The present invention also includes a gene of interest operably-linked to a second gene containing a signal sequence. An example of such an embodiment is wherein the gene of interest is operably-linked to the ovalbumin gene that contains an ovalbumin signal sequence. Other signal sequences that can be included in the transposon-based vectors include, but are not limited to the ovotransferrin and lysozyme signal sequences. In one embodiment, the signal sequence is an ovalbumin signal sequence including a sequence shown in SEQ ID NO:74. In another embodiment, the signal sequence is a shortened ovalbumin signal sequence including a sequence shown in SEQ ID NO:75 or SEQ ID NO:76.

As also used herein, the term "targeting sequence" refers to an amino acid sequence, or the polynucleotide sequence encoding the amino acid sequence, which amino acid sequence is recognized by a receptor located on the exterior of a cell. Binding of the receptor to the targeting sequence results in uptake of the protein or peptide operably-linked to the targeting sequence by the cell. One example of a targeting sequence is a vitellogenin targeting sequence that is recognized by a vitellogenin receptor (or the low density lipoprotein receptor) on the exterior of an oocyte. In one embodiment, the vitellogenin targeting sequence includes the polynucleotide sequence of SEQ ID NO:77. In another embodiment, the vitellogenin targeting sequence includes all or part of the vitellogenin gene. Other targeting

sequences include VLDL and Apo E, which are also capable of binding the vitellogenin receptor. Since the ApoE protein is not endogenously expressed in birds, its presence may be used advantageously to identify birds carrying the transposon-based vectors of the present invention.

#### 5 Genes of Interest

10

15

20

25

30

The genes of interest in the polynucleotide cassette can be any gene, and preferably are genes that encode proteins or portions of multimeric proteins that may be used in vaccines. Such genes may be coding regions for desired epitopes in infectious organisms, including but not limited to viral, bacterial, protozoal, fungal, and parasitic organisms. Such genes are known to one of ordinary skill in the art and may be obtained from GenBank and from scientific and patent literature. A gene of interest may contain modifications of the codons for the first several N-terminal amino acids of the gene of interest, wherein the third base of each codon is changed to an A or a T without changing the corresponding amino acid.

In one embodiment, the genes of interest are antibody genes or portions of antibody genes. Figure 2 shows a schematic drawing of a polynucleotide cassette containing an antibody heavy chain and an antibody light chain as two genes of interest. Antibodies used in or encoded by the polynucleotide cassettes of the present invention include, but are not limited to, IgG, IgM, IgA, IgD, IgE, IgY, lambda chains, kappa chains, bi-specific antibodies, and fragments thereof; scFv fragments, Fc fragments, and Fab fragments as well as dimeric, trimeric and oligomeric forms of antibody fragments. Desired antibodies include, but are not limited to, naturally occurring antibodies, human antibodies, humanized antibodies, autoantibodies and hybrid antibodies. Preferred antibodies include those useful as vaccines against Genes encoding modified versions of naturally occurring selected diseases. antibodies or fragments thereof and genes encoding artificially designed antibodies or fragments thereof may be incorporated into the transposon-based vectors of the present invention. Desired antibodies also include antibodies with the ability to bind specific ligands, for example, antibodies against proteins associated with diseases, particularly infectious diseases. Accordingly, the present invention encompasses a polynucleotide cassette as described herein containing one or more genes encoding a heavy immunoglobulin (Ig) chain and a light Ig chain.

5

10

15

20

25

30

Antibodies that may be produced using the present invention include, but are not limited to, antibodies for providing passive immunity to an animal or a human against an infectious disease or a toxic agent. The antibodies prepared using the methods of the present invention may also be designed to possess specific labels that may be detected through means known to one of ordinary skill in the art. For example, antibodies may be labeled with a fluorescent label attached that may be detected following exposure to specific wavelengths. Such labeled antibodies may be primary antibodies directed to a specific antigen, for example, rhodamine-labeled rabbit anti-growth hormone, or may be labeled secondary antibodies, such as fluorescein-labeled goat-anti chicken IgG. Such labeled antibodies are known to one of ordinary skill in the art. The antibodies may also be designed to possess specific sequences useful for purification through means known to one of ordinary skill in the art. Labels useful for attachment to antibodies are also known to one of ordinary skill in the art. Some of these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety. Proteins and multimeric proteins made according to the present invention may also be labeled with these methods. Antibodies, proteins and multimeric proteins produced with the present invention may be used as laboratory reagents for numerous applications including radioimmunoassay, western blots, dot blots, ELISA, immunoaffinity columns and other procedures requiring antibodies as known to one of ordinary skill in the art. Such antibodies include primary antibodies, secondary antibodies and tertiary antibodies, which may be labeled or unlabeled.

Additional antibodies that may be made with the practice of the present invention include, but are not limited to, primary antibodies, secondary antibodies, designer antibodies, anti-protein antibodies, anti-peptide antibodies, anti-DNA antibodies, anti-RNA antibodies, antibodies against non-natural antigens, anti-venom antibodies, antibodies directed against epitopes associated with infectious disease, including, anti-viral, anti-bacterial, anti-protozoal, anti-fungal, anti-parasitic, anti-receptor, anti-lipid, anti-phospholipid, anti-growth factor, anti-cytokine, anti-monokine, anti-idiotype, and anti-accessory (presentation) protein antibodies. Antibodies made with the present invention, as well as light chains or heavy chains, may also be used to inhibit enzyme activity.

5

10

15

Antibodies that may be produced using the present invention include, but are not limited to, antibodies made against the following proteins: Bovine  $\gamma$ -Globulin, Serum; Bovine IgG, Plasma; Chicken  $\gamma$ -Globulin, Serum; Human  $\gamma$ -Globulin, Serum; Human IgA, Plasma; Human IgA1, Myeloma; Human IgA2, Myeloma; Human IgA2, Plasma; Human IgD, Plasma; Human IgE, Myeloma; Human IgG, Plasma; Human IgG, Fab Fragment, Plasma; Human IgG, F(ab')2 Fragment, Plasma; Human IgG, Fc Fragment, Plasma; Human IgG1, Myeloma; Human IgG2, Myeloma; Human IgG3, Myeloma; Human IgG4, Myeloma; Human IgM, Plasma; Human Immunoglobulin, Light Chain  $\kappa$ , Urine; Human Immunoglobulin, Light Chains  $\kappa$  and  $\lambda$ , Plasma; Mouse  $\gamma$ -Globulin, Serum; Mouse IgG, Serum; Mouse IgM, Myeloma; Rabbit  $\gamma$ -Globulin, Serum; Rabbit IgG, Plasma; and Rat  $\gamma$ -Globulin, Serum. In one embodiment, the transposon-based vector comprises the coding sequence of light and heavy chains of a murine monoclonal antibody that shows specificity for human seminoprotein (GenBank Accession numbers AY129006 and AY129304 for the light and heavy chains, respectively).

A further non-limiting list of antibodies that recognize other antibodies and that may be produced using the present invention is as follows: Anti-Chicken IgG, heavy (H) & light (L) Chain Specific (Sheep); Anti-Goat γ-Globulin (Donkey); Anti-Goat IgG, Fc Fragment Specific (Rabbit); Anti-Guinea Pig γ-Globulin (Goat); Anti-Human Ig, Light Chain, Τγρε κ Specific; Anti-Human Ig, Light Chain, Τγρε λ 20 Specific; Anti-Human IgA, α-Chain Specific (Goat); Anti-Human IgA, Fab Fragment Specific; Anti-Human IgA, Fc Fragment Specific; Anti-Human IgA, Secretory; Anti-Human IgE, ε-Chain Specific (Goat); Anti-Human IgE, Fc Fragment Specific; Anti-Human IgG, Fc Fragment Specific (Goat); Anti-Human IgG,  $\gamma$ -Chain Specific (Goat); Anti-Human IgG, Fc Fragment Specific; Anti-Human IgG, Fd Fragment Specific; 25 Anti-Human IgG, H & L Chain Specific (Goat); Anti-Human IgG1, Fc Fragment Specific; Anti-Human IgG2, Fc Fragment Specific; Anti-Human IgG2, Fd Fragment Specific; Anti-Human IgG3, Hinge Specific; Anti-Human IgG4, Fc Fragment Specific; Anti-Human IgM, Fc Fragment Specific; Anti-Human IgM, μ-Chain Specific; Anti-Mouse IgE, ε-Chain Specific; Anti-Mouse γ-Globulin (Goat); Anti-Mouse IgG, γ-30 Chain Specific (Goat); Anti-Mouse IgG, \( \gamma\)-Chain Specific (Goat) F(ab')2 Fragment; Anti-Mouse IgG, H & L Chain Specific (Goat); Anti-Mouse IgM, µ-Chain Specific (Goat); Anti-Mouse IgM, H & L Chain Specific (Goat); Anti-Rabbit γ-Globulin

(Goat); Anti-Rabbit IgG, Fc Fragment Specific (Goat); Anti-Rabbit IgG, H & L Chain Specific (Goat); Anti-Rat  $\gamma$ -Globulin (Goat); Anti-Rat IgG, H & L Chain Specific; Anti-Rhesus Monkey  $\gamma$ -Globulin (Goat); and, Anti-Sheep IgG, H & L Chain Specific.

Antibodies may be made against peptides such as p146 (SEQ ID NO:78 amino acid sequence, SEQ ID NO:79, nucleotide sequence), and other lytic peptides or peptides associated with disease, particularly infectious disease.

5

10

15

20

25

30

The following is yet another non-limiting of antibodies that can be produced by the methods of present invention: abciximab (ReoPro), abciximab anti-platelet aggregation monoclonal antibody, anti-CD11a (hu1124), anti-CD18 antibody, anti-CD20 antibody, anti-cytomegalovirus (CMV) antibody, anti-digoxin antibody, antihepatitis B antibody, anti-HER-2 antibody, anti-idiotype antibody to GD3 glycolipid, anti-IgE antibody, anti-IL-2R antibody, antimetastatic cancer antibody (mAb 17-1A), anti-rabies antibody, anti-respiratory syncytial virus (RSV) antibody, anti-Rh antibody, anti-TCR, anti-TNF antibody, anti-VEGF antibody and Fab fragment thereof, rattlesnake venom antibody, black widow spider venom antibody, coral snake venom antibody, antibody against very late antigen-4 (VLA-4), C225 humanized antibody to EGF receptor, chimeric (human & mouse) antibody against TNFa, antibody directed against GPIIb/IIIa receptor on human platelets, gamma globulin, anti-hepatitis B immunoglobulin, human anti-D immunoglobulin, human antibodies against S aureus, human tetanus immunoglobulin, humanized antibody against the epidermal growth receptor-2, humanized antibody against the  $\alpha$  subunit of the interleukin-2 receptor, humanized antibody CTLA4IG, humanized antibody to the IL-2 R α-chain, humanized anti-CD40-ligand monoclonal antibody (5c8), humanized mAb against the epidermal growth receptor-2, humanized mAb to rous sarcoma virus, humanized recombinant antibody (IgG1k) against respiratory syncytial virus (RSV), lymphocyte immunoglobulin (anti-thymocyte antibody), lymphocyte immunoglobulin, mAb against factor VII, MDX-210 bi-specific antibody against HER-2, MDX-22, MDX-220 bi-specific antibody against TAG-72 on tumors, MDX-33 antibody to FcγR1 receptor, MDX-447 bi-specific antibody against EGF receptor, MDX-447 bispecific humanized antibody to EGF receptor, MDX-RA immunotoxin (ricin A linked) antibody, Medi-507 antibody (humanized form of BTI-322) against CD2 receptor on T-cells, monoclonal antibody LDP-02, muromonab-CD3(OKT3)

5

10

15

20

25

30

antibody, OKT3 ("muromomab-CD3") antibody, PRO 542 antibody, ReoPro ("abciximab") antibody, and TNF-IgG fusion protein.

Antibodies may also be made, in accordance with the present invention, against the following non-limiting list of organisms associated with disease and against peptides and/or proteins or other epitopes from these organisms: Anthrax (Bacillus anthracis), Bacillus anthracis (anthrax), Botulism (Clostridium botulinum toxin), Brucella species (brucellosis), Brucellosis (Brucella species), Burkholderia mallei (glanders), Burkholderia pseudomallei (melioidosis), Chlamydia psittaci (psittacosis), Cholera (Vibrio cholerae), Clostridium botulinum toxin (botulism), Clostridium perfringens (Epsilon toxin), Coxiella burnetii (Q fever), E. coli O157:H7 (Escherichia coli), Emerging infectious diseases such as Nipah virus and hantavirus, Epsilon toxin of Clostridium perfringens, Escherichia coli O157:H7 (E. coli), Food safety threats (e.g., Salmonella species, Escherichia coli O157:H7, Shigella), Francisella tularensis (tularemia), Glanders (Burkholderia mallei), Melioidosis (Burkholderia pseudomallei), Plague (Yersinia pestis), Psittacosis (Chlamydia psittaci), Q fever (Coxiella burnetii), Ricin toxin from Ricinus communis (castor beans), Rickettsia prowazekii (typhus fever), Salmonella species (salmonellosis), Salmonella Typhi (typhoid fever), Salmonellosis (Salmonella species), Shigella (shigellosis), Shigellosis (Shigella), Smallpox (variola major), Staphylococcal enterotoxin B, Tularemia (Francisella tularensis), Typhoid fever (Salmonella Typhi), Typhus fever (Rickettsia prowazekii), Variola major (smallpox), Vibrio cholerae (cholera), Viral encephalitis (alphaviruses [e.g., Venezuelan equine encephalitis, eastern equine encephalitis, western equine encephalitis]), Viral hemorrhagic fevers (filoviruses [e.g., Ebola, Marburg] and arenaviruses [e.g., Lassa, Machupo]), Water safety threats (e.g., Vibrio cholerae, Cryptosporidium parvum), and Yersinia pestis (plague).

Another non-limiting list of the antibodies that may be produced using the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. (www.phoenixpeptide.com; 530 Harbor Boulevard, Belmont, CA), Peninsula Labs San Carlos CA, SIGMA, St.Louis, MO www.sigma-aldrich.com, Cappel ICN, Irvine, California, www.icnbiomed.com, and Calbiochem, La Jolla, California, www.calbiochem.com, which are all incorporated herein by reference in their entirety. The polynucleotide sequences encoding these antibodies

5

10

15

20

25

30

may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the antibody polynucleotide sequence by choosing the codons that encode for each amino acid in the desired antibody.

As an example of protein production using the method of the present invention which may be applied to the production of proteins useful in vaccines, the gene of interest is a proinsulin gene and the desired molecule is insulin. Proinsulin consists of three parts: a C-peptide and two strands of amino acids (the alpha and beta chains) that later become linked together to form the insulin molecule. In these embodiments, proinsulin is expressed in the oviduct tubular gland cells and then deposited in the egg white. One example of a proinsulin polynucleotide sequence is shown in SEQ ID NO:80, wherein the C-peptide cleavage site spans from Arg at position 31 to Arg at position 65.

Further included in the present invention are genes of interet that encode proteins and peptides synthesized by the immune system including those synthesized by the thymus, lymph nodes, spleen, and the gastrointestinal associated lymph tissues (GALT) system. The immune system proteins and peptides proteins that can be made in transgenic animals using the polynucleotide cassettes of the present invention include, but are not limited to, alpha-interferon, beta-interferon, gamma-interferon, alpha-interferon A, alpha-interferon 1, G-CSF, GM-CSF, interlukin-1 (IL-1), IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, TNF-α, and TNF-β. Other cytokines included in the present invention include cardiotrophin, stromal cell derived factor, macrophage derived chemokine (MDC), melanoma growth stimulatory activity (MGSA), macrophage inflammatory proteins 1 alpha (MIP-1 alpha), 2, 3 alpha, 3 beta, 4 and 5.

Genes encoding lytic peptides such as p146 are also included in the genes of interest of the present invention. In one embodiment, the p146 peptide comprises an amino acid sequence of SEQ ID NO:78. The present invention also encompasses a polynucleotide cassette comprising a p146 nucleic acid having a sequence of SEQ ID NO:79.

Enzymes are another class of proteins that may be encoded by the polynucleotide cassettes of the present invention. Such enzymes include but are not limited to adenosine deaminase, alpha-galactosidase, cellulase, collagenase, dnaseI,

hyaluronidase, lactase, L-asparaginase, pancreatin, papain, streptokinase B, subtilisin, superoxide dismutase, thrombin, trypsin, urokinase, fibrinolysin, glucocerebrosidase and plasminogen activator. In some embodiments wherein the enzyme could have deleterious effects, additional amino acids and a protease cleavage site are added to the carboxy end of the enzyme of interest in order to prevent expression of a functional enzyme. Subsequent digestion of the enzyme with a protease results in activation of the enzyme. Preferred are enzymes associate with disease processes and the function of infectious organisms.

A non-limiting list of the peptides and proteins that may be encoded by the polynucleotide cassettes of the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. (www.phoenixpeptide.com; 530 Harbor Boulevard, Belmont, CA), Peninsula Labs (San Carlos CA), SIGMA, (St.Louis, MO www.sigma-aldrich.com), Cappel ICN (Irvine, California, www.icnbiomed.com), and Calbiochem (La Jolla, California, www.calbiochem.com). The polynucleotide sequences encoding these proteins and peptides of interest may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the polynucleotide sequence to be incorporated into the genome by choosing the codons that encode for each amino acid in the desired protein or peptide.

Other desired proteins that may be encoded by the polynucleotide cassettes of the present invention include bacitracin, polymixin b, vancomycin, cyclosporine, anti-RSV antibody, alpha-1 antitrypsin (AAT), anti-cytomegalovirus antibody, antihepatitis antibody, anti-inhibitor coagulant complex, anti-rabies antibody, anti-Rh(D) antibody, adenosine deaminase, anti-digoxin antibody, antivenin crotalidae (rattlesnake venom antibody), antivenin latrodectus (black widow spider venom antibody), antivenin micrurus (coral snake venom antibody), aprotinin, corticotropin (ACTH), diphtheria antitoxin, lymphocyte immune globulin (anti-thymocyte antibody), protamine, thyrotropin, capreomycin, α-galactosidase, gramicidin, streptokinase, tetanus toxoid, tyrothricin, IGF-1, proteins of varicella vaccine, anti-TNF antibody, anti-IL-2r antibody, anti-HER-2 antibody, OKT3 ("muromonab-CD3") antibody, TNF-IgG fusion protein, ReoPro ("abciximab") antibody, ACTH fragment 1-24, desmopressin, gonadotropin-releasing hormone, histrelin, leuprolide, lypressin, nafarelin, peptide that binds GPIIb/GPIIIa on platelets (integrilin),

5

10

15

20

25

30

goserelin, capreomycin, colistin, anti-respiratory syncytial virus, lymphocyte immune globulin (Thymoglovin, Atgam), panorex, alpha-antitrypsin, botulinin, lung surfactant protein, tumor necrosis receptor-IgG fusion protein (enbrel), gonadorelin, proteins of influenza vaccine, proteins of rotavirus vaccine, proteins of haemophilus b conjugate vaccine, proteins of poliovirus vaccine, proteins of pneumococcal conjugate vaccine, proteins of meningococcal C vaccine, proteins of influenza vaccine, megakaryocyte growth and development factor (MGDF), neuroimmunophilin ligand-A (NIL-A), brain-derived neurotrophic factor (BDNF), glial cell line-derived neurotrophic factor (GDNF), leptin (native), leptin B, leptin C, IL-1RA (interleukin-1RA), R-568, novel erythropoiesis-stimulating protein (NESP), humanized mAb to rous sarcoma virus (MEDI-493), glutamyl-tryptophan dipeptide IM862, LFA-3TIP immunosuppressive, humanized anti-CD40-ligand monoclonal antibody (5c8), gelsonin enzyme, tissue factor pathway inhibitor (TFPI), proteins of meningitis B vaccine, antimetastatic cancer antibody (mAb 17-1A), chimeric (human & mouse) mAb against TNFα, mAb against factor VII, relaxin, capreomycin, glycopeptide (LY333328), recombinant human activated protein C (rhAPC), humanized mAb against the epidermal growth receptor-2, altepase, anti-CD20 antigen, C2B8 antibody, insulin-like growth factor-1, atrial natriuretic peptide (anaritide), tenectaplase, anti-CD11a antibody (hu 1124), anti-CD18 antibody, mAb LDP-02, anti-VEGF antibody, fab fragment of anti-VEGF Ab, APO2 ligand (tumor necrosis factor-related apoptosis-inducing ligand), rTGF-β (transforming growth factor- $\beta$ ), alpha-antitrypsin, ananain (a pineapple enzyme), humanized mAb CTLA4IG, PRO 542 (mAb), D2E7 (mAb), calf intestine alkaline phosphatase, α-L-iduronidase, α-L-galactosidase (humanglutamic acid decarboxylase, acid sphingomyelinase, bone morphogenetic protein-2 (rhBMP-2), proteins of HIV vaccine, T cell receptor (TCR) peptide vaccine, TCR peptides, V beta 3 and V beta 13.1. (IR502), (IR501), BI 1050/1272 mAb against very late antigen-4 (VLA-4), C225 humanized mAb to EGF receptor, anti-idiotype antibody to GD3 glycolipid, antibacterial peptide against H. pylori, MDX-447 bispecific humanized mAb to EGF receptor, anti-cytomegalovirus (CMV), Medi-491 B19 parvovirus vaccine, humanized recombinant mAb (IgG1k) against respiratory syncytial virus (RSV), urinary tract infection vaccine (against "pili" on Escherechia coli strains), proteins of lyme disease vaccine against B. burgdorferi protein (DbpA), proteins of Medi-501 human papilloma virus-11 vaccine (HPV), Streptococcus pneumoniae vaccine, Medi-507

mAb (humanized form of BTI-322) against CD2 receptor on T-cells, MDX-33 mAb to Fc $\gamma$ R1 receptor, MDX-RA immunotoxin (ricin A linked) mAb, MDX-210 bispecific mAb against HER-2, MDX-447 bi-specific mAb against EGF receptor, MDX-22, MDX-220 bi-specific mAb against TAG-72 on tumors, colony-stimulating factor (CSF) (molgramostim), humanized mAb to the IL-2 R  $\alpha$ -chain (basiliximab), mAb to IgE (IGE 025A), myelin basic protein-altered peptide (MSP771A), humanized mAb against the epidermal growth receptor-2, humanized mAb against the  $\alpha$  subunit of the interleukin-2 receptor, low molecular weight heparin, anti-hemophillic factor, and bactericidal/permeability-increasing protein (r-BPI).

The following paragraph describes specific examples of epitopes and antibodies associated with specific diseases that may be made as the proteins, multimeric proteins of the present invention. Such epitopes may also be used in antibody production. Anthrax: PA (protective antigen) mediates entry of toxin's lethal factor (LF) into cell, therefore PA is binding domain. There are at least three different binding sites on PA where LF can bind. Site I epitope IG3PA63, site II 2D3PA, 2D5PA, 10D2PA, and site III 3B6PA, 14B7PA, 10E10PA63. Botulism: BoNT (botulism neurotoxin), binding domains within neurotoxin A (BoNT/A) derived from segments of heavy chain and encoded proteins: H455-661, H1150-1289. Accession number X73844. BoNT/A in infants: Accession number X73423.

Tularemia: mAbs against LPS of francisella tularensis: FT14, FT2F11. OMP specific mAbs react to IgGl subclass of *F. tularensis*. Yersenia pestis: causes pneumonic and bubonic plague. LCrV: accession number M26405 (*Y. pestis*), LCrV: accession number M57893 (*Y. pseudotuberculosis*). Caf-1 encodes capsular antigen fraction 1 (F1) accession number X61996. Shigella: epitopes 2C32E6, 4D64B9 (reactive to *S. flexneri* and *S. boydii*). 5E45D8 – *S. flexneri*. 4B33D10, 1B52F10 – all species of Shieglla. Shigella monoclonal antibodies 2C32E6, 4D64B9, 5E45D8 (reactive with S. Felxneri). 4B33D10 and 1B52F1 (species of Shigella) and 1B52F10 all strains (see Rahman & Stimson, Hybridoma, 20:85-90, 2001). See Venkatesan et al., PNAS 85:9317-21 (1988) accession number J04117 S. flexnari). IpaD between amino acid residues 14 and 44 of Shigella flexnari. IpaD is against the unique region of this protein exposed on the surface of the pathogen (Turbyfill et al., Infect. Immunity 66:1999-2006 (1998). See Muller-Loennies et al., J. Biol. Chem. 278: 25618-27 (2003) for epitopes related to Shigella. LcrV, Yersinia pestis (accession.

5

10

15

20

25

30

number M26405), Yersinia pseudotuberculosis (accessionnumber M57893, Yersinia enterocolitica. Fusion of IcrV and the structural gene for protein A. One protective epitope resides internally between amino acids 168 and 275 (Motin et al., Infect. Immunity 62:41920-201 (1994). For additional epitopes on Yersinia pastis see accession number M26405 LcrV (Price et al., J. Bacteriol. 171: 5646-5653 (1989). For additional epitopes on Yersinia pastis see cafl encoding cafl capsular antigen fraction 1 (F1) accession number X61996 in Galyov et al., FEBS Lett. 277: 230-2 (1990). 12G12 SLP-M common specific epitope of the Brucella S-LPS. S-LPS of B. melitensis Rev1, S-LPS smooth lipopolysaccharide, (Weynants et al., Clin. Diagn. Lab. Immunol. 3:309-14 (1996). Smooth Brucella species - two lipopolysacchiaride (LPSs (A and M) produce monoclonal antibodies to the specific A- or M-LPS epitopes. 12AE6, and Mab 33.1.5, s-LPSA=12AE6, LPSM=33.1.5, BmE10-5, 12G12 O side chain = BRU-38 (Douglas et al., J. Clin. Microbiol. 26: 1353-6 (1988). BmE10-5 Mantigen of Brucella melitensis 16M, B. melitensis, B. abortus, B. suis and B. neotomae (Vizcaino et al., Res. Microbiol. 143: 513-8 (1992). BRU-38, O side chain of B. abortus (Roop et al., J. Clin. Microbiol. 25: 2090-3 (1987). The OMPspecific monoclonal antibodies Bhatti et al., Hybridoma 12: 197-202 (1993). Mabs FT14 and FT2F11, Francisella tularensis (Fulop et al., J. Clin. Microbiol. 29: 1407-12 (1991).

Preferred proteins and multimeric proteins to be made with the present invention include those associated with disease processes and those associated with antigenic epitopes of infectious organisms.

The proteins and multimeric proteins made using the present invention may be labeled using labels and techniques known to one of ordinary skill in the art. Some of these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety. Some of these labels may be genetically engineered into the polynucleotide sequence for the expression of the selected multimeric protein. The peptides and proteins may also have label-incorporation "handles" incorporated to allow labeling of an otherwise difficult or impossible to label multimeric protein.

It is to be understood that the various classes of desired peptides and proteins, as well as specific peptides and proteins described in this section may be modified as

described below by inserting selected codons for desired amino acid substitutions into the gene incorporated into the transgenic animal.

The present invention further encompasses the use of inhibitory molecules to inhibit endogenous (i.e., non-vector) protein production. These inhibitory molecules include antisense nucleic acids, siRNA and inhibitory proteins. Preferred inhibitory molecules are those associated with causation or inhibition of disease processes. In a preferred embodiment, the endogenous protein whose expression is inhibited is an egg white protein including, but not limited to ovalbumin, ovotransferrin, and ovomucin ovomucoid, ovoinhibitor, cystatin, ovostatin, lysozyme, ovoglobulin G2, ovoglobulin G3, avidin, and thiamin binding protein. In one embodiment, a polynucleotide cassette containing an ovalbumin DNA sequence, that upon transcription forms a double stranded RNA molecule, is transfected into an animal such as a bird and the bird's production of endogenous ovalbumin protein is reduced by the interference RNA mechanism (RNAi). In other embodiments, a polynucleotide cassette encodes an inhibitory RNA molecule that inhibits the expression of more than one egg white protein. Additionally, inducible knockouts or knockdowns of the endogenous protein may be created to achieve a reduction or inhibition of endogenous protein production. Endogenous egg white production can be inhibited in an avian at any time, but is preferably inhibited preceding, or immediately preceding, the harvest of eggs.

## Modified Desired Proteins and Peptides

5

10

15

20

25

30

The present invention may be used for the production of proteins and multimeric proteins. "Proteins", "peptides," "polypeptides" and "oligopeptides" are chains of amino acids (typically L-amino acids) whose alpha carbons are linked through peptide bonds formed by a condensation reaction between the carboxyl group of the alpha carbon of one amino acid and the amino group of the alpha carbon of another amino acid. The terminal amino acid at one end of the chain (i.e., the amino terminal) has a free amino group, while the terminal amino acid at the other end of the chain (i.e., the carboxy terminal) has a free carboxyl group. As such, the term "amino terminus" (abbreviated N-terminus) refers to the free alpha-amino group on the amino acid at the amino terminal of the protein, or to the alpha-amino group (imino group when participating in a peptide bond) of an amino acid at any other location within the protein. Similarly, the term "carboxy terminus" (abbreviated C-terminus) refers to

the free carboxyl group on the amino acid at the carboxy terminus of a protein, or to the carboxyl group of an amino acid at any other location within the protein.

Typically, the amino acids making up a protein are numbered in order, starting at the amino terminal and increasing in the direction toward the carboxy terminal of the protein. Thus, when one amino acid is said to "follow" another, that amino acid is positioned closer to the carboxy terminal of the protein than the preceding amino acid.

The term "residue" is used herein to refer to an amino acid (D or L) or an amino acid mimetic that is incorporated into a protein by an amide bond. As such, the amino acid may be a naturally occurring amino acid or, unless otherwise limited, may encompass known analogs of natural amino acids that function in a manner similar to the naturally occurring amino acids (i.e., amino acid mimetics). Moreover, an amide bond mimetic includes peptide backbone modifications well known to those skilled in the art.

Furthermore, one of skill will recognize that, as mentioned above, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than about 5%, more typically less than about 1%) in an encoded sequence are conservatively modified variations where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (O);
- 25 4) Arginine (R), Lysine (K);

5

10

15

20

30

- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

A conservative substitution is a substitution in which the substituting amino acid (naturally occurring or modified) is structurally related to the amino acid being substituted, i.e., has about the same size and electronic properties as the amino acid being substituted. Thus, the substituting amino acid would have the same or a similar functional group in the side chain as the original amino acid. A "conservative substitution" also refers to utilizing a substituting amino acid which is identical to the

5

10

15

20

25

amino acid being substituted except that a functional group in the side chain is protected with a suitable protecting group.

Suitable protecting groups are described in Green and Wuts, "Protecting Groups in Organic Synthesis", John Wiley and Sons, Chapters 5 and 7, 1991, the teachings of which are incorporated herein by reference. Preferred protecting groups are those which facilitate transport of the peptide through membranes, for example, by reducing the hydrophilicity and increasing the lipophilicity of the peptide, and which can be cleaved, either by hydrolysis or enzymatically (Ditter et al., 1968. J. Pharm. Sci. 57:783; Ditter et al., 1968. J. Pharm. Sci. 57:828; Ditter et al., 1969. J. Pharm. Sci. 58:557; King et al., 1987. Biochemistry 26:2294; Lindberg et al., 1989. Drug Metabolism and Disposition 17:311; Tunek et al., 1988. Biochem. Pharm. 37:3867; Anderson et al., 1985 Arch. Biochem. Biophys. 239:538; and Singhal et al., 1987. FASEB J. 1:220). Suitable hydroxyl protecting groups include ester, carbonate and carbamate protecting groups. Suitable amine protecting groups include acyl groups and alkoxy or aryloxy carbonyl groups, as described above for N-terminal protecting groups. Suitable carboxylic acid protecting groups include aliphatic, benzyl and aryl esters, as described below for C-terminal protecting groups. In one embodiment, the carboxylic acid group in the side chain of one or more glutamic acid or aspartic acid residues in a peptide of the present invention is protected, preferably as a methyl, ethyl, benzyl or substituted benzyl ester, more preferably as a benzyl ester.

Provided below are groups of naturally occurring and modified amino acids in which each amino acid in a group has similar electronic and steric properties. Thus, a conservative substitution can be made by substituting an amino acid with another amino acid from the same group. It is to be understood that these groups are non-limiting, i.e. that there are additional modified amino acids which could be included in each group.

- Group I includes leucine, isoleucine, valine, methionine and modified amino acids having the following side chains: ethyl, n-propyl n-butyl. Preferably, Group I includes leucine, isoleucine, valine and methionine.
- 30 Group II includes glycine, alanine, valine and a modified amino acid having an ethyl side chain. Preferably, Group II includes glycine and alanine.
  - Group III includes phenylalanine, phenylglycine, tyrosine, tryptophan, cyclohexylmethyl glycine, and modified amino residues having substituted

benzyl or phenyl side chains. Preferred substituents include one or more of the following: halogen, methyl, ethyl, nitro, —NH<sub>2</sub>, methoxy, ethoxy and — CN. Preferably, Group III includes phenylalanine, tyrosine and tryptophan.

Group IV includes glutamic acid, aspartic acid, a substituted or unsubstituted aliphatic, aromatic or benzylic ester of glutamic or aspartic acid (e.g., methyl, ethyl, n-propyl iso-propyl, cyclohexyl, benzyl or substituted benzyl), glutamine, asparagine, —CO—NH— alkylated glutamine or asparagines (e.g., methyl, ethyl, n-propyl and iso-propyl) and modified amino acids having the side chain —(CH<sub>2</sub>)<sub>3</sub>—COOH, an ester thereof (substituted or unsubstituted aliphatic, aromatic or benzylic ester), an amide thereof and a substituted or unsubstituted N-alkylated amide thereof. Preferably, Group IV includes glutamic acid, aspartic acid, methyl aspartate, ethyl aspartate, benzyl aspartate and methyl glutamate, ethyl glutamate and benzyl glutamate, glutamine and asparagine.

Group V includes histidine, lysine, ornithine, arginine, N-nitroarginine, β-cycloarginine, γ-hydroxyarginine, N-amidinocitruline and 2-amino-4-guanidinobutanoic acid, homologs of lysine, homologs of arginine and homologs of ornithine. Preferably, Group V includes histidine, lysine, arginine and ornithine. A homolog of an amino acid includes from 1 to about
 3 additional or subtracted methylene units in the side chain.

Group VI includes serine, threonine, cysteine and modified amino acids having C1-C5 straight or branched alkyl side chains substituted with —OH or —SH, for example, —CH<sub>2</sub>CH<sub>2</sub>OH, —CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>OH or -CH<sub>2</sub>CH<sub>2</sub>OHCH<sub>3</sub>. Preferably, Group VI includes serine, cysteine or threonine.

25

30

In another aspect, suitable substitutions for amino acid residues include "severe" substitutions. A "severe substitution" is a substitution in which the substituting amino acid (naturally occurring or modified) has significantly different size and/or electronic properties compared with the amino acid being substituted. Thus, the side chain of the substituting amino acid can be significantly larger (or smaller) than the side chain of the amino acid being substituted and/or can have functional groups with significantly different electronic properties than the amino acid being substituted. Examples of severe substitutions of this type include the substitution of phenylalanine or cyclohexylmethyl glycine for alanine, isoleucine for

glycine, a D amino acid for the corresponding L amino acid, or —NH—CH[(—CH<sub>2</sub>)<sub>5</sub>—COOH]—CO— for aspartic acid. Alternatively, a functional group may be added to the side chain, deleted from the side chain or exchanged with another functional group. Examples of severe substitutions of this type include adding of valine, leucine or isoleucine, exchanging the carboxylic acid in the side chain of aspartic acid or glutamic acid with an amine, or deleting the amine group in the side chain of lysine or ornithine. In yet another alternative, the side chain of the substituting amino acid can have significantly different steric and electronic properties that the functional group of the amino acid being substituted. Examples of such modifications include tryptophan for glycine, lysine for aspartic acid and — (CH<sub>2</sub>)<sub>4</sub>COOH for the side chain of serine. These examples are not meant to be limiting.

In another embodiment, for example in the synthesis of a peptide 26 amino acids in length, the individual amino acids may be substituted according in the following manner:

AA<sub>1</sub> is serine, glycine, alanine, cysteine or threonine;

AA2 is alanine, threonine, glycine, cysteine or serine;

AA<sub>3</sub> is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine,  $\beta$ -cycloarginine,  $\gamma$ -hydroxyarginine, N-amidinocitruline or 2-amino-4-

20 guanidinobutanoic acid;

5

10

15

AA4 is proline, leucine, valine, isoleucine or methionine;

AA5 is tryptophan, alanine, phenylalanine, tyrosine or glycine;

AA6 is serine, glycine, alanine, cysteine or threonine;

AA<sub>7</sub> is proline, leucine, valine, isoleucine or methionine;

25 AA<sub>8</sub> is alanine, threonine, glycine, cysteine or serine;

AA9 is alanine, threonine, glycine, cysteine or serine;

AA<sub>10</sub> is leucine, isoleucine, methionine or valine;

AA11 is serine, glycine, alanine, cysteine or threonine;

AA<sub>12</sub> is leucine, isoleucine, methionine or valine;

30 AA<sub>13</sub> is leucine, isoleucine, methionine or valine;

AA<sub>14</sub> is glutamine, glutamic acid, aspartic acid, asparagine, or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

 $AA_{15}$  is arginine, N-nitroarginine,  $\beta$ -cycloarginine,  $\gamma$ -hydroxy-arginine, N-amidinocitruline or 2-amino-4-guanidino-butanoic acid

AA<sub>16</sub> is proline, leucine, valine, isoleucine or methionine;

AA<sub>17</sub> is serine, glycine, alanine, cysteine or threonine;

5 AA<sub>18</sub> is glutamic acid, aspartic acid, asparagine, glutamine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

AA<sub>19</sub> is aspartic acid, asparagine, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

10 AA<sub>20</sub> is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine, β-cycloarginine, γ-hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;

AA21 is alanine, threonine, glycine, cysteine or serine;

AA<sub>22</sub> is alanine, threonine, glycine, cysteine or serine;

20

25

30

15 AA<sub>23</sub> is histidine, serine, threonine, cysteine, lysine or ornithine;

AA<sub>24</sub> is threonine, aspartic acid, serine, glutamic acid or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

AA<sub>25</sub> is asparagine, aspartic acid,, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid; and

AA<sub>26</sub> is cysteine, histidine, serine, threonine, lysine or ornithine.

It is to be understood that these amino acid substitutions may be made for longer or shorter peptides than the 26 mer in the preceding example above, and for proteins.

In one embodiment of the present invention, codons for the first several N-terminal amino acids of the transposase are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. It is preferable that between approximately 1 and 20, more preferably 3 and 15, and most preferably between 4 and 12 of the first N-terminal codons of the gene of interest are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. In one embodiment, the first ten N-terminal codons of the gene of interest are modified in this manner.

When several desired proteins, protein fragments or peptides are encoded in the gene of interest to be incorporated into the genome, as with the multivalent multimeric proteins, one of skill in the art will appreciate that the proteins, protein fragments or peptides may be separated by a spacer molecule such as, for example, a peptide, consisting of one or more amino acids. Generally, the spacer will have no specific biological activity other than to join the desired proteins, protein fragments or peptides together, or to preserve some minimum distance or other spatial relationship between them. However, the constituent amino acids of the spacer may be selected to influence some property of the molecule such as the folding, net charge, or hydrophobicity. The spacer may also be contained within a nucleotide sequence with a purification handle or be flanked by proteolytic cleavage sites.

Such polypeptide spacers may have from about 1 to about 100 amino acids, preferably 3 to 20 amino acids, and more preferably 4-15 amino acids. The spacers in a polypeptide are independently chosen, but are preferably all the same. The spacers should allow for flexibility of movement in space and are therefore typically rich in small amino acids, for example, glycine, serine, proline or alanine. Preferably, peptide spacers contain at least 60%, more preferably at least 80% glycine or alanine. In addition, peptide spacers generally have little or no biological and antigenic activity. Preferred spacers are (Gly-Pro-Gly-Gly)<sub>x</sub> (SEQ ID NO:81) and (Gly<sub>4</sub>-Ser)<sub>y</sub>, wherein x is an integer from about 3 to about 9 and y is an integer from about 1 to about 8. Specific examples of suitable spacers include

(Gly-Pro-Gly-Gly)3

5

10

15

20

30

SEQ ID NO:82 Gly Pro Gly Gly Gly Pro Gly Gly Pro Gly Gly (Gly<sub>4</sub>-Ser)<sub>3</sub>

25 SEQ ID NO:83 Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser Gly Gly Ser or (Gly4-Ser)4

One example of a multivalent multimeric protein containin a spacer is leutinizing hormone (LH), normally made as separate alpha and beta chains, made as a single polypeptide as described in Galet et. al., Mol. Cell Endocrinology, 2001, 174 (1-2):31-40. Production of a multimeric protein may thus be simplified using a spacer sequence that may or may not contain cleavage sites. In the case of an

immunoglobulin, for example, a heavy and light chain may be synthesized as a single polypeptide using a spacer sequence with protease sites native to the transgenic animal so as to make, upon processing, a heavy and light chain combination in close association, facilitating the addition of a similar heavy and light chain to produce the native immunoglobulin. In this model, the removal of the spacer sequence may or may not be required. Other multimeric proteins may be made in bioengineered organisms in a similar fashion.

5

10

15

20

25

30

Nucleotide sequences encoding for the production of residues which may be useful in purification of the expressed recombinant protein may also be built into the vector. Such sequences are known in the art and include the glutathione binding domain from glutathione S-transferase, polylysine, hexa-histidine or other cationic amino acids, thioredoxin, hemagglutinin antigen and maltose binding protein.

Additionally, nucleotide sequences may be inserted into the gene of interest to be incorporated so that the protein or peptide can also include from one to about six amino acids that create signals for proteolytic cleavage. In this manner, if a gene is designed to make one or more peptides or proteins of interest in the transgenic animal, specific nucleotide sequences encoding for amino acids recognized by enzymes may be incorporated into the gene to facilitate cleavage of the large protein or peptide sequence into desired peptides or proteins or both. For example, nucleotides encoding a proteolytic cleavage site can be introduced into the gene of interest so that a signal sequence can be cleaved from a protein or peptide encoded by the gene of interest. Nucleotide sequences encoding other amino acid sequences which display pH sensitivity, chemical sensitivity or photolability may also be added to the vector to facilitate separation of the signal sequence from the peptide or protein of interest.

Proteolytic cleavage sites include cleavage sites recognized by exopeptidases such as carboxypeptidase A, carboxypeptidase B, aminopeptidase I, and dipeptidylaminopeptidase; endopeptidases such as trypsin, V8-protease, enterokinase, factor Xa, collagenase, endoproteinase, subtilisin, and thombin; and proteases such as Protease 3C IgA protease (Igase) Rhinovirus 3C(preScission)protease. Chemical cleavage sites are also included in the defintion of cleavage site as used herein. Chemical cleavage sites include, but are not limited to, site cleaved by cyanogen bromide, hydroxylamine, formic acid, and acetic acid. Self-splicing cleavage sites such as inteins are also included in the present invention.

In some embodiments, one or more cleavage sites are incorporated into a polynucleotide cassette containing multiple genes of interest. Figure 4 depicts one example of a polynucleotide cassette containing two genes of interest containing a cleavage site between them. The genes of interest may encode different proteins or peptides, the same protein or peptide, or modified versions of the same protein or peptide. While Figure 4 shows a polynucleotide cassette containing two genes of interest, the present invention encompasses a polynucleotide cassette containing any number of genes of interest. The cleavage site located between the genes of interest can encode any amino acid sequence that is cleaved by any means. As mentioned above, the cleavage site can encode an amino acid sequence cleaved by a protease, a chemical reaction, can be a photolabile site, or can be a pro polynucleotide.

The present invention includes a polynucleotide cassette that encodes a repetitive polypeptide chain in which two or more peptides, polypeptides or proteins, designated as P in the structural formulae presented below, are each separated by a peptide spacer or cleavage site designated as B. A polypeptide multivalent ligand, also called a multivalent protein, is a form of a multimeric protein encoded by the polynucleotide cassettes of the present invention, and is represented by structural formulae (I, II and III). Each peptide or protein is connected to another peptide or protein through a peptide bond, to a linker group, to a spacer, or to a cleavage site. Each peptide, polypeptide or protein may be the same or different and each linker, spacer, cleavage site or covalent bond is independently chosen.

A "polypeptide multivalent protein" is a multiple repeat polypeptide chain in which two or more peptides P are each separated by a peptide linker group, a spacer or a cleavage site. A polypeptide multivalent ligand is represented by structural formulae II and III.

$$I - B - (L - P)n$$

5

10

15

20

25

30

wherein B is a peptide spacer or cleavage site, n is an integer from 2 to about 20, each L is a covalent bond, a linking group or cleavage site which may be present or absent, and each P is a peptide having from about 4 to about 200 amino acid residues.

wherein m is an integer from 0 to about 20.

III 
$$Pa - (B)n - Pa$$

wherein n is an integer from 1 to 20, preferably 2 to 10, more preferably 3 to 7, further wherein a is 1.

Other examples of multivalent proteins include the following:

$$P_y$$
  $L_x$   $B_n$   $L_x$   $P_y$ 

10

15

20

25

30

$$V$$
  $P_y$   $B$   $P_y$ 

In the preceding structural formulae IV and V of polypeptide multivalent ligands encoded by a polynucleotide cassette of the present invention, each P is a peptide having from about 4 to about 200 amino acid residues, y is 1, x is an integer from 1 to 3, and n is an integer from 1 to 20, preferably 2 to 10, more preferably 3 to 7. Each B is a peptide spacer or cleavage site comprised of at least 2 amino acids or a cleavage site. Each peptide P and each B are independently chosen and may be the same or different.

Suitable linkers (L) are groups that can connect peptides and proteins to each other. In one example, the linker is an oligopeptide of from about 1 to about 10 amino acids consisting of amino acids with inert side chains. Suitable oligopeptides include polyglycine, polyserine, polyproline, polyalanine and oligopeptides consisting of alanyl and/or serinyl and/or prolinyl and/or glycyl amino acid residues. m in structural formula II is an integer from 0 to about 20.

The peptides, polypeptides and proteins in a multivalent protein can be connected to each other by covalent bonds, linker groups, spacers, cleavage groups or a combination thereof. The linking groups can be the same or different.

A polypeptide spacer shown in structural formula (II) is a peptide having from about 5 to about 40 amino acid residues. The spacers in a polypeptide multivalent ligand are independently chosen, and may be the same or different. The spacers should allow for flexibility of movement in space for the flanking peptides. polypeptides and proteins P, and are therefore typically rich in small amino acids, for

example, glycine, serine, proline or alanine. Preferably, peptide spacers contain at least 60%, more preferably at least 80% glycine or alanine. In addition, peptide spacers generally have little or no biological and antigenic activity. Preferred spacers are (Gly-Pro-Gly-Gly)<sub>x</sub> (SEQ ID NO:81) and (Gly<sub>4</sub>-Ser)<sub>y</sub>, wherein x is an integer from about 3 to about 9 and y is an integer from about 1 to about 8. Specific examples of suitable spacers include (Gly<sub>4</sub>-Ser)<sub>3</sub> (SEQ ID NO:82). Spacers can also include from one to about four amino acids that create a signal for proteolytic cleavage.

5

10

15

20

25

30

In another embodiment of the present invention, a TAG sequence is linked to a gene of interest. The TAG sequence serves three purposes: 1) it allows free rotation of the peptide or protein to be isolated so there is no interference from the native protein or signal sequence, i.e. vitellogenin, 2) it provides a "purification handle" to isolate the protein using affinity purification, and 3) it includes a cleavage site to remove the desired protein from the signal and purification sequences. Accordingly, as used herein, a TAG sequence includes a spacer sequence, a purification handle and a cleavage site. The spacer sequences in the TAG proteins contain one or more repeats shown in SEQ ID NO:85. A preferred spacer sequence comprises the sequence provided in SEQ ID NO:86. One example of a purification handle is the gp41 hairpin loop from HIV I. Exemplary gp41 polynucleotide and polypeptide sequences are provided in SEQ ID NO:87 and SEQ ID NO:88, respectively. However, it should be understood that any antigenic region, or otherwise associative regions such as avidin/biotin, may be used as a purification handle, including any antigenic region of gp41. Preferred purification handles are those that elicit highly specific antibodies. Additionally, the cleavage site can be any protein cleavage site known to one of ordinary skill in the art and includes an enterokinase cleavage site comprising the Asp Asp Asp Asp Lys sequence (SEQ ID NO:89) and a furin cleavage In one embodiment of the present invention, the TAG sequence comprises a polynucleotide sequence of SEQ ID NO:90.

## Methods of Administering Polynucleotide Cassettes

In addition to the polynucleotide cassettes described above, the present invention also includes methods of administering the polynucleotide cassettes to an animal, methods of producing a transgenic animal wherein a gene of interest is incorporated into the germline of the animal and methods of producing a transgenic

animal wherein a gene of interest is incorporated into cells other than the germline cells of the animal. The polynucleotide cassettes may reside in any vector or delivery solution when administered or may be naked DNA. In one embodiment, a transposon-based vector containing the polynucleotide cassette between two insertion sequences recognized by a transposase is administered to an animal. The polynucleotide cassettes of the present invention may be administered to an animal via any method known to those of skill in the art, including, but not limited to, intraembryonic, intratesticular, intraoviduct, intraovarian, into the duct system of the mammary gland, intraperitoneal, intraarterial, intravenous, topical, oral, nasal, and pronuclear injection methods of administration, or any combination thereof. The polynucleotide cassettes may also be administered within the lumen of an organ, into an organ, into a body cavity, into the cerebrospinal fluid, through the urinary system, through the genitourinary system, through the reproductive system, or through any route to reach the desired cells.

The polynucleotide cassettes may be delivered through the vascular system to be distributed to the cells supplied by that vessel. For example, the compositions may be placed in the artery supplying the ovary or supplying the fallopian tube to transfect cells in those tissues. In this manner, follicles could be transfected to create a germline transgenic animal. Alternatively, supplying the compositions through the artery leading to the oviduct would preferably transfect the tubular gland and epithelial cells. Such transfected cells could manufacture a desired protein or peptide for deposition in the egg white. Administration of the compositions through the portal vein would target uptake and transformation of hepatic cells. Administration through the urethra and into the bladder would target the transitional epithelium of the bladder. Administration through the vagina and cervix would target the lining of the uterus and the epithelial cells of the fallopian tube. Administration through the internal mammary artery or through the duct system of the mammary gland would transfect secretory cells of the lactating mammary gland to perform a desired function, such as to synthesize and secrete a desired protein or peptide into the milk.

The polynucleotide cassettes may be administered in a single administration, multiple administrations, continuously, or intermittently. The polynucleotide cassettes may be administered by injection, via a catheter, an osmotic mini-pump or any other method. In some embodiments, a polynucleotide cassette is administered to

5

10

15

20

25

30

an animal in multiple administrations, each administration containing the polynucleotide cassette and a different transfecting reagent.

In a preferred embodiment, the animal is an egg-laying animal, and more preferably, an avian. In one embodiment, between approximately 1 and 150 µg, 1 and 100 µg, 1 and 50 µg, preferably between 1 and 20 µg, and more preferably between 5 and 10 µgof a transposon-based vector containing the polynucleotide cassette is administered to the oviduct of a bird. In a chicken, it is preferred that between approximately 1 and 100 µg, or 5 and 50 µg are administered. In a quail, it is preferred that between approximately 5 and 10 µg are administered. Optimal ranges depending upon the type of bird and the bird's stage of sexual maturity. Intraoviduct administration of the transposon-based vectors of the present invention result in a PCR positive signal in the oviduct tissue, whereas intravascular administration results in a PCR positive signal in the liver. In other embodiments, the polynucleotide cassettes is administration may also be combined with any methods for facilitating transfection, including without limitation, electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

The transposon-based vectors may be administered to the animal at any point during the lifetime of the animal, however, it is preferable that the vectors are administered prior to the animal reaching sexual maturity. The transposon-based vectors are preferably administered to a chicken oviduct between approximately 14 and 16 weeks of age and to a quail oviduct between approximately 5 and 10 weeks of age, more preferably 5 and 8 weeks of age, and most preferably between 5 and 6 weeks of age, when standard poultry rearing practices are used. The vectors may be administered at earlier ages when exogenous hormones are used to induce early sexual maturation in the bird. In some embodiments, the transposon-based vector is administered to an animal's oviduct following an increase in proliferation of the Such an increase in oviduct epithelial cells and/or the tubular gland cells. proliferation normally follows an influx of reproductive hormones in the area of the oviduct. When the animal is an avian, the transposon-based vector is administered to the avian's oviduct following an increase in proliferation of the oviduct epithelial cells and before the avian begins to produce egg white constituents.

The present invention also includes a method of intraembryonic administration of a transposon-based vector containing a polynucleotide cassette to an avian embryo comprising the following steps: 1) incubating an egg on its side at room temperature for two hours to allow the embryo contained therein to move to top dead center (TDC); 2) drilling a hole through the shell without penetrating the underlying shell membrane; 3) injecting the embryo with the transposon-based vector in solution; 4) sealing the hole in the egg; and 5) placing the egg in an incubator for hatching. Administration of the transposon-based vector can occur anytime between immediately after egg lay (when the embryo is at Stage X) and hatching. Preferably, the transposon-based vector is administered between 1 and 7 days after egg lay, more preferably between 1 and 2 days after egg lay. The transposon-based vectors may be introduced into the embryo in amounts ranging from about 5.0 µg to 10 pg, preferably 1.0 µg to 100 pg. Additionally, the transposon-based vector solution volume may be between approximately 1 µl to 75 µl in quail and between approximately 1 µl to 500 µl in chicken.

The present invention also includes a method of intratesticular administration of a transposon-based vector containing a polynucleotide cassette including injecting a bird with a composition comprising the transposon-based vector, an appropriate carrier and an appropriate transfection reagent. In one embodiment, the bird is injected before sexual maturity, preferably between approximately 4-14 weeks, more preferably between approximately 6-14 weeks and most preferably between 8-12 weeks old. In another embodiment, a mature bird is injected with a transposon-based vector an appropriate carrier and an appropriate transfection reagent. The mature bird may be any type of bird, but in one example the mature bird is a quail.

A bird is preferably injected prior to the development of the blood-testis barrier, which thereby facilitates entry of the transposon-based vector into the seminiferous tubules and transfection of the spermatogonia or other germline cells. At and between the ages of 4, 6, 8, 10, 12, and 14 weeks, it is believed that the testes of chickens are likely to be most receptive to transfection. In this age range, the blood/testis barrier has not yet formed, and there is a relatively high number of spermatogonia relative to the numbers of other cell types, e.g., spermatids, etc. See J. Kumaran et al., 1949. Poultry Sci., 29:511-520. See also E. Oakberg, 1956. Am. J. Anatomy, 99:507-515; and P. Kluin et al., 1984. Anat. Embryol., 169:73-78.

5

10

15

20

25

30

The transposon-based vectors may be introduced into a testis in an amount ranging from about 0.1  $\mu$ g to 10  $\mu$ g, preferably 1  $\mu$ g to 10  $\mu$ g, more preferably 3  $\mu$ g to 10  $\mu$ g. In a quail, about 5  $\mu$ g is a preferred amount. In a chicken, about 5  $\mu$ g to 10  $\mu$ g per testis is preferred. These amounts of vector DNA may be injected in one dose or multiple doses and at one site or multiple sites in the testis. In a preferred embodiment, the vector DNA is administered at multiple sites in a single testis, both testes being injected in this manner. In one embodiment, injection is spread over three injection sites: one at each end of the testis, and one in the middle. Additionally, the transposon-based vector solution volume may be between approximately 1  $\mu$ l to 75  $\mu$ l in quail and between approximately 1  $\mu$ l to 500  $\mu$ l in chicken. In a preferred embodiment, the transposon-based vector solution volume may be between approximately 20  $\mu$ l to 60  $\mu$ l in quail and between approximately 50  $\mu$ l to 250  $\mu$ l in chicken. Both the amount of vector DNA and the total volume injected into each testis may be determined based upon the age and size of the bird.

According to the present invention, the polynucleotide cassette is administered in conjunction with an acceptable carrier and/or transfection reagent. Acceptable carriers include, but are not limited to, water, saline, Hanks Balanced Salt Solution (HBSS), Tris-EDTA (TE) and lyotropic liquid crystals. Transfection reagents commonly known to one of ordinary skill in the art that may be employed include, but are not limited to, the following: cationic lipid transfection reagents, cationic lipid mixtures, polyamine reagents, liposomes and combinations thereof; SUPERFECT®, Cytofectene, BioPORTER®, GenePORTER®, NeuroPORTER®, and perfectin from Gene Therapy Systems; lipofectamine, cellfectin, DMRIE-C oligofectamine, and PLUS reagent from InVitrogen; Xtreme gene, fugene, DOSPER and DOTAP from Roche; Lipotaxi and Genejammer from Strategene; and Escort from SIGMA. In one embodiment, the transfection reagent is SUPERFECT®. The ratio of DNA to transfection reagent may vary based upon the method of administration. In one embodiment, a transposon-based vector containing a polynucleotide cassette is administered intratesticularly and the ratio of DNA to transfection reagent can be from 1:1.5 to 1:15, preferably 1:2 to 1:10, all expressed as wt/vol. Transfection may also be accomplished using other means known to one of ordinary skill in the art, including without limitation electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

Depending upon the cell or tissue type targeted for transfection, the form of the transposon-based vector may be important. Plasmids harvested from bacteria are generally closed circular supercoiled molecules, and this is the preferred state of a vector for gene delivery because of the ease of preparation. In some instances, transposase expression and insertion may be more efficient in a relaxed, closed circular configuration or in a linear configuration. In still other instances, a purified transposase protein may be co-injected with a transposon-based vector containing the gene of interest for more immediate insertion. This could be accomplished by using a transfection reagent complexed with both the purified transposase protein and the transposon-based vector.

# Testing for and Breeding Animals Carrying the Transgene

5

10

15

20

25

30

Following administration of a polynucleotide cassette to an animal, DNA is extracted from the animal to confirm integration of the genes of interest. Advantages provided by the present invention include the high rates of integration, or incorporation, and transcription of the gene of interest when administered to a bird via an intraoviduct or intraovary route (including intraarterial administrations to arteries leading to the oviduct or ovary) and contained within a transposon-based vector.

Actual frequencies of integration can be estimated both by comparative strength of the PCR signal, and by histological evaluation of the tissues by quantitative PCR. Another method for estimating the rate of transgene insertion is the so-called primed in situ hybridization technique (PRINS). This method determines not only which cells carry a transgene of interest, but also into which chromosome the gene has inserted, and even what portion of the chromosome. Briefly, labeled primers are annealed to chromosome spreads (affixed to glass slides) through one round of PCR, and the slides are then developed through normal in situ hybridization procedures. This technique combines the best features of in situ PCR and fluorescence in situ hybridization (FISH) to provide distinct chromosome location and copy number of the gene in question. The 28s rRNA gene will be used as a positive control for spermatogonia to confirm that the technique is functioning properly. Using different fluorescent labels for the transgene and the 28s gene causes cells containing a transgene to fluoresce with two different colored tags.

Breeding experiments may also be conducted to determine if germline transmission of the transgene has occurred. In a general bird breeding experiment

performed according to the present invention, each male bird is exposed to 2-3 different adult female birds for 3-4 days each. This procedure is continued with different females for a total period of 6-12 weeks. Eggs are collected daily for up to 14 days after the last exposure to the transgenic male, and each egg is incubated in a standard incubator. The resulting embryos are examined for transgene presence at day 3 or 4 using PCR.

5

10

15

20

25

30

Any male producing a transgenic embryo is bred to additional females. Eggs from these females are incubated, hatched, and the chicks tested for the exogenous DNA. Any embryos that die are necropsied and examined directly for the transgene or protein encoded by the transgene, either by fluorescence or PCR. The offspring that hatch and are found to be positive for the exogenous DNA are raised to maturity. These birds are bred to produce further generations of transgenic birds, to verify efficiency of the transgenic procedure and the stable incorporation of the transgene into the germ line. The resulting embryos are examined for transgene presence at day 3 or 4 using PCR.

It is to be understood that the above procedure can be modified to suit animals other than birds and that selective breeding techniques may be performed to amplify gene copy numbers and protein output.

# Production of Desired Proteins and Multimeric Proteins in Egg White

In one embodiment, a transposon-based vector containing a polynucleotide cassette of the present invention may be administered to a bird for production of desired proteins or peptides in the egg white. These trassposon-based vectors preferably contain one or more of an ovalbumin promoter, an ovomucoid promoter, an ovalbumin signal sequence and an ovomucoid signal sequence. Oviduct-specific ovalbumin promoters are described in B. O'Malley et al., 1987. EMBO J., vol. 6, pp. 2305-12; A. Qiu et al., 1994. Proc. Nat. Acad. Sci. (USA), vol. 91, pp. 4451-4455; D. Monroe et al., 2000. Biochim. Biophys. Acta, 1517 (1):27-32; H. Park et al., 2000. Biochem., 39:8537-8545; and T. Muramatsu et al., 1996. Poult. Avian Biol. Rev., 6:107-123.

# Production of Desired Proteins and Multimeric Proteins in Egg Yolk

The present invention is particularly advantageous for production of recombinant peptides and proteins of low solubility in the egg yolk. Such proteins include, but are not limited to, membrane-associated or membrane-bound proteins,

lipophilic compounds; attachment factors, receptors, and components of second messenger transduction machinery. Low solubility peptides and proteins are particularly challenging to produce using conventional recombinant protein production techniques (cell and tissue cultures) because they aggregate in water-based, hydrophilic environments. Such aggregation necessitates denaturation and refolding of the recombinantly-produced proteins, which may deleteriously affect their structure and function. Moreover, even highly soluble recombinant peptides and proteins may precipitate and require denaturation and renaturation when produced in sufficiently high amounts in recombinant protein production systems. The present invention provides an advantageous resolution of the problem of protein and peptide solubility during production of large amounts of recombinant proteins.

5

10

15

20

25

30

In one embodiment of the present invention, deposition of a desired protein into the egg yolk is accomplished by attaching a sequence encoding a protein capable of binding to the yolk vitellogenin receptor to a gene of interest that encodes a desired protein. This polynucleotide cassette can be used for the receptor-mediated uptake of the desired protein by the oocytes. In a preferred embodiment, the sequence ensuring the binding to the vitellogenin receptor is a targeting sequence of a vitellogenin protein. The invention encompasses various vitellogenin proteins and their targeting sequences. In a preferred embodiment, a chicken vitellogenin protein targeting sequence is used, however, due to the high degree of conservation among vitellogenin protein sequences and known cross-species reactivity of vitellogenin targeting sequences with their egg-yolk receptors, other vitellogenin targeting sequences can be substituted. One example of a construct for use in the transposon-based vectors of the present invention and for deposition of an insulin protein in an egg yolk is a transposon-based vector containing a vitellogenin promoter, a vitellogenin targeting sequence, a TAG sequence, a pro-insulin sequence and a synthetic polyA sequence. The present invention includes, but is not limited to, vitellogenin targeting sequences residing in the N-terminal domain of vitellogenin, particularly in lipovitellin I. In one embodiment, the vitellogenin targeting sequence contains the polynucleotide sequence of SEQ ID NO:77.

In a preferred embodiment, the transposon-based vector contains a transposase gene operably-linked to a constitutive promoter and a gene of interest operably-linked to a liver-specific promoter and a vitellogenin targeting sequence.

### Isolation and Purification of Desired Proteins and Multimeric Proteins

5

10

15

20

25

30

For large-scale production of protein, an animal breeding stock that is homozygous for the transgene is preferred. Such homozygous individuals are obtained and identified through, for example, standard animal breeding procedures or PCR protocols.

Once expressed, peptides, polypeptides and proteins can be purified according to standard procedures known to one of ordinary skill in the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis, high performance liquid chromatography, immunoprecipitation and the like. Substantially pure compositions of about 50 to 99% homogeneity are preferred, and 80 to 95% or greater homogeneity are most preferred for use as therapeutic agents.

In one embodiment of the present invention, the animal in which the desired protein is produced is an egg-laying animal. In a preferred embodiment of the present invention, the animal is an avian and a desired peptide, polypeptide or protein is isolated from an egg white. Egg white containing the exogenous protein or peptide is separated from the yolk and other egg constituents on an industrial scale by any of a variety of methods known in the egg industry. See, e.g., W. Stadelman et al. (Eds.), Egg Science & Technology, Haworth Press, Binghamton, NY (1995). Isolation of the exogenous peptide or protein from the other egg white constituents is accomplished by any of a number of polypeptide isolation and purification methods well known to one of ordinary skill in the art. These techniques include, for example, chromatographic methods such as gel permeation, ion exchange, affinity separation, metal chelation, HPLC, hydrophobic interaction chromatography, and the like, either alone or in combination. Such techniques are known to one of ordinary skill in the art of protein purification. Another means that may be used for isolation or purification, either in lieu of or in addition to chromatographic separation methods, includes Successful isolation and purification is confirmed by standard electrophoresis. analytic techniques, including HPLC, mass spectroscopy, and spectrophotometry. These separation methods are often facilitated if the first step in the separation is the removal of the endogenous ovalbumin fraction of egg white, as doing so will reduce the total protein content to be further purified by about 50%.

To facilitate or enable purification of a desired protein or peptide, the polynucleotide cassettes may include one or more additional epitopes or domains.

Such epitopes or domains include DNA sequences encoding enzymatic, chemical or photolabile cleavage sites including, but not limited to, an enterokinase cleavage site; the glutathione binding domain from glutathione S-transferase; polylysine; hexahistidine or other cationic amino acids, and sites cleaved by cyanogen bromide, hydroxylamine, formic acid, and acetic acid; thioredoxin; hemagglutinin antigen; maltose binding protein; a fragment of gp41 from HIV; and other purification epitopes or domains commonly known to one of skill in the art. Other proteolytic cleavage sites that may be included in the polynucleotide cassettes are cleavage sites recognized by exopeptidases such as carboxypeptidase A, carboxypeptidase B, aminopeptidase I, and dipeptidylaminopeptidase; endopeptidases such as trypsin, V8-protease, enterokinase, factor Xa, collagenase, endoproteinase, subtilisin, and thombin; and proteases such as Protease 3C IgA protease (Igase) Rhinovirus 3C(preScission)protease. Self-splicing cleavage sites such as inteins may also be included in the polynucleotide cassettes of the present invention.

10

15

20

25

30

In one representative embodiment, purification of desired proteins from egg white utilizes the antigenicity of the ovalbumin carrier protein and particular attributes of a TAG linker sequence that spans ovalbumin and the desired protein. The TAG sequence is particularly useful in this process because it contains 1) a highly antigenic epitope, a fragment of gp41 from HIV, allowing for stringent affinity purification, and, 2) a recognition site for the protease enterokinase immediately juxtaposed to the desired protein. In a preferred embodiment, the TAG sequence comprises approximately 50 amino acids. A representative TAG sequence is provided below.

The underlined sequences were taken from the hairpin loop domain of HIV gp-41 (SEQ ID NO:87). Sequences in italics represent the cleavage site for enterokinase (SEQ ID NO:89). The spacer sequence upstream of the loop domain was made from repeats of (Pro Ala Asp Asp Ala) (SEQ ID NO:85) to provide free rotation and promote surface availability of the hairpin loop from the ovalbumin carrier protein.

Isolation and purification of a desired protein is performed as follows:

1. Enrichment of the egg white protein fraction containing ovalbumin and the transgenic ovalbumin-TAG-desired protein.

- 2. Size exclusion chromatography to isolate only those proteins within a narrow range of molecular weights (a further enrichment of step 1).
- Ovalbumin affinity chromatography. Highly specific antibodies to ovalbumin
  will eliminate virtually all extraneous egg white proteins except ovalbumin
  and the transgenic ovalbumin-TAG-desired protein.

10

15

20

25

30

- 4. gp41 affinity chromatography using anti-gp41 antibodies. Stringent application of this step will result in virtually pure transgenic ovalbumin-TAG-desired protein.
- 5. Cleavage of the transgene product can be accomplished in at least one of two ways:
  - a. The transgenic ovalbumin-TAG-desired protein is left attached to the gp41 affinity resin (beads) from step 4 and the protease enterokinase is added. This liberates the transgene target protein from the gp41 affinity resin while the ovalbumin-TAG sequence is retained. Separation by centrifugation (in a batch process) or flow through (in a column purification), leaves the desired protein together with enterokinase in solution. Enterokinase is recovered and reused.
  - b. Alternatively, enterokinase is immobilized on resin (beads) by the addition of poly-lysine moieties to a non-catalytic area of the protease. The transgenic ovalbumin-TAG-desired protein eluted from the affinity column of step 4 is then applied to the protease resin. Protease action cleaves the ovalbumin-TAG sequence from the desired protein and leaves both entities in solution. The immobilized enterokinase resin is recharged and reused.
  - c. The choice of these alternatives is made depending upon the size and chemical composition of the transgene target protein.
- 6. A final separation of either of these two (5a or 5b) protein mixtures is made using size exclusion, or enterokinase affinity chromatography. This step allows for desalting, buffer exchange and/or polishing, as needed.

Cleavage of the transgene product (ovalbumin-TAG-desired protein) by enterokinase, then, results in two products: ovalbumin-TAG and the desired protein.

More specific methods for isolation using the TAG label is provided in the Examples. Some desired proteins may require additions or modifications of the above-described approach as known to one of ordinary skill in the art. The method is scaleable from the laboratory bench to pilot and production facility largely because the techniques applied are well documented in each of these settings.

It is believed that a typical chicken egg produced by a transgenic animal of the present invention will contain at least 0.001 mg, from about 0.001 to 1.0 mg, or from about 0.001 to 100.0 mg of exogenous protein, peptide or polypeptide, in addition to the normal constituents of egg white (or possibly replacing a small fraction of the latter). One of skill in the art will recognize that after biological expression or purification, the desired proteins, fragments thereof and peptides may possess a conformation substantially different than the native conformations of the proteins, fragments thereof and peptides. In this case, it is often necessary to denature and reduce protein and then to cause the protein to re-fold into the preferred conformation. Methods of reducing and denaturing proteins and inducing re-folding are well known to those of skill in the art.

### Production of Proteins and Multimeric Proteins in Milk

5

10

15

20

25

30

In addition to methods of producing eggs containing transgenic proteins or peptides, the present invention encompasses methods for the production of milk containing transgenic proteins or peptides. These methods include the administration of a transposon-based vector described above to a mammal, for example through administration into the mammary duct system in order to access epithelial cells. In another embodiment, the transposon-based vector may be administered into arteries supplying the mammary gland. In a further embodiment germline transformed animals may have the transposon-based vector stably incorporated into the mammary In one embodiment, the transposon-based vector contains a epithelial cells. transposase operably-linked to a constitutive promoter and a gene of interest operably-linked to mammary specific promoter. Genes of interest can include, but are not limited to antiviral and antibacterial proteins, immunoglobulins, antibodies and proteins associated with disease and infectious organisms. Milk may be processed for isolation of proteins, multimeric proteins and antibodies using techniques known to one of skill in the art of processing milk and purifying proteins and antibodies from milk.

# Administration of the Vaccines Produced with the Present Invention

The proteins, polypeptides, peptides and antibodies made with the present invention for use in vaccines may be administered to animals or humans using techniques known to one of ordinary skill in the art. Clinical practitioners of ordinary skill in the in the veterinary and human medical fields are familiar with methods, routes, volumes and schedules of vaccine administration, as well as various adjuvants that may be optionally employed to engender or amplify an immune response.

The following examples will serve to further illustrate the present invention without, at the same time, however, constituting any limitation thereof. On the contrary, it is to be clearly understood that resort may be had to various embodiments, modifications and equivalents thereof which, after reading the description herein, may suggest themselves to those skilled in the art without departing from the spirit of the invention.

15 EXAMPLE 1

5

10

20

25

30

Preparation of Transposon-Based Vector pTnMod

A vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells, given below as SEQ ID NO:57. The vector of SEQ ID NO:57, termed pTnMod, was constructed and its sequence verified.

This vector employed a cytomegalovirus (CMV) promoter. A modified Kozak sequence (ACCATG) (SEQ ID NO:5) was added to the promoter. The nucleotide in the wobble position in nucleotide triplet codons encoding the first 10 amino acids of transposase was changed to an adenine (A) or thymine (T), which did not alter the amino acid encoded by this codon. Two stop codons were added and a synthetic polyA was used to provide a strong termination sequence. This vector uses a promoter designed to be active soon after entering the cell (without any induction) to increase the likelihood of stable integration. The additional stop codons and synthetic polyA insures proper termination without read through to potential genes downstream.

The first step in constructing this vector was to modify the transposase to have the desired changes. Modifications to the transposase were accomplished with the primers High Efficiency forward primer (Hef) Altered transposase (ATS)-Hef 5' ATCTCGAGACCATGTGTGAACTTGATATTTTACATGATTCTCTTTACC 3'

5

10

15

20

25

30

(SEQ ID NO:91) and Altered transposase- High efficiency reverse primer (Her) 5' GATTGATCATAATTTCCCCAAAGCGTAACC 3' (SEQ ID NO:92, a reverse complement primer). In the 5' forward primer ATS-Hef, the sequence CTCGAG (SEQ ID NO:93) is the recognition site for the restriction enzyme Xho I, which permits directional cloning of the amplified gene. The sequence ACCATG (SEQ ID NO:5) contains the Kozak sequence and start codon for the transposase and the underlined bases represent changes in the wobble position to an A or T of codons for the first 10 amino acids (without changing the amino acid coded by the codon). Primer ATS-Her (SEQ ID NO:92) contains an additional stop codon TAA in addition to native stop codon TGA and adds a Bcl I restriction site, TGATCA (SEQ ID NO:94), to allow directional cloning. These primers were used in a PCR reaction with pTnLac (p defines plasmid, tn defines transposon, and lac defines the beta fragment of the lactose gene, which contains a multiple cloning site) as the template for the transposase and a FailSafe<sup>TM</sup> PCR System (which includes enzyme, buffers, dNTP's, MgCl<sub>2</sub> and PCR Enhancer; Epicentre Technologies, Madison, WI). Amplified PCR product was electrophoresed on a 1% agarose gel, stained with ethidium bromide, and visualized on an ultraviolet transilluminator. corresponding to the expected size was excised from the gel and purified from the agarose using a Zymo Clean Gel Recovery Kit (Zymo Research, Orange, CA). Purified DNA was digested with restriction enzymes Xho I (5') and Bcl I (3') (New England Biolabs, Beverly, MA) according to the manufacturer's protocol. Digested DNA was purified from restriction enzymes using a Zymo DNA Clean and Concentrator kit (Zymo Research).

Plasmid gWhiz (Gene Therapy Systems, San Diego, CA) was digested with restriction enzymes Sal I and BamH I (New England Biolabs), which are compatible with Xho I and Bcl I, but destroy the restriction sites. Digested gWhiz was separated on an agarose gel, the desired band excised and purified as described above. Cutting the vector in this manner facilitated directional cloning of the modified transposase (mATS) between the CMV promoter and synthetic polyA.

To insert the mATS between the CMV promoter and synthetic polyA in gWhiz, a Stratagene T4 Ligase Kit (Stratagene, Inc. La Jolla, CA) was used and the ligation set up according to the manufacturer's protocol. Ligated product was transformed into *E. coli* Top10 competent cells (Invitrogen Life Technologies,

5

10

15

20

25

30

Carlsbad, CA) using chemical transformation according to Invitrogen's protocol. Transformed bacteria were incubated in 1 ml of SOC (GIBCO BRL, CAT# 15544-042) medium for 1 hour at 37° C before being spread to LB (Luria-Bertani media (broth or agar)) plates supplemented with 100 µg/ml ampicillin (LB/amp plates). These plates were incubated overnight at 37° C and resulting colonies picked to LB/amp broth for overnight growth at 37° C. Plasmid DNA was isolated using a modified alkaline lysis protocol (Sambrook et al., 1989), electrophoresed on a 1% agarose gel, and visualized on a U.V. transilluminator after ethidium bromide staining. Colonies producing a plasmid of the expected size (approximately 6.4 kbp) were cultured in at least 250 ml of LB/amp broth and plasmid DNA harvested using a Oiagen Maxi-Prep Kit (column purification) according to the manufacturer's protocol (Qiagen, Inc., Chatsworth, CA). Column purified DNA was used as template for sequencing to verify the changes made in the transposase were the desired changes and no further changes or mutations occurred due to PCR amplification. For sequencing, Perkin-Elmer's Big Dye Sequencing Kit was used. All samples were sent to the Gene Probes and Expression Laboratory (LSU School of Veterinary Medicine) for sequencing on a Perkin-Elmer Model 377 Automated Sequencer.

Once a clone was identified that contained the desired mATS in the correct orientation, primers CMVf-NgoM IV (5' TTGCCGGCATCAGATTGGCTAT (SEQ ID NO:95); underlined bases denote NgoM IV recognition site) and Syn-polyA-BstE II (5' AGAGGTCACCGGGTCAATTCTTCAGCACCTGGTA (SEQ ID NO:96); underlined bases denote BstE II recognition site) were used to PCR amplify the entire CMV promoter, mATS, and synthetic polyA for cloning upstream of the transposon in pTnLac. The PCR was conducted with FailSafe<sup>TM</sup> as described above, purified using the Zymo Clean and Concentrator kit, the ends digested with NgoM IV and BstE II (New England Biolabs), purified with the Zymo kit again and cloned upstream of the transposon in pTnLac as described below.

Plasmid pTnLac was digested with NgoM IV and BstE II to remove the ptac promoter and transposase and the fragments separated on an agarose gel. The band corresponding to the vector and transposon was excised, purified from the agarose, and dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs) to prevent self-annealing. The enzyme was removed from the vector using a Zymo DNA Clean and Concentrator-5. The purified vector and CMVp/mATS/polyA

were ligated together using a Stratagene T4 Ligase Kit and transformed into E. coli as described above.

Colonies resulting from this transformation were screened (mini-preps) as describe above and clones that were the correct size were verified by DNA sequence analysis as described above. The vector was given the name pTnMod (SEQ ID NO:57) and includes the following components:

5

10

15

20

25

30

Base pairs 1-130 are a remainder of F1(-) on from pBluescriptll sk(-) (Stratagene), corresponding to base pairs 1-130 of pBluescriptll sk(-).

Base pairs 131 - 132 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 133 -1777 are the CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems), corresponding to bp 229-1873 of pGWiz. The CMV promoter was modified by the addition of an ACC sequence upstream of ATG.

Base pairs 1778-1779 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 1780 - 2987 are the coding sequence for the transposase, modified from Tn10 (GenBank accession J01829) by optimizing codons for stability of the transposase mRNA and for the expression of protein. More specifically, in each of the codons for the first ten amino acids of the transposase, G or C was changed to A or T when such a substitution would not alter the amino acid that was encoded.

Base pairs 2988-2993 are two engineered stop codons.

Base pair 2994 is a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 2995 - 3410 are a synthetic polyA sequence taken from the pGWiz vector (Gene Therapy Systems), corresponding to bp 1922-2337 of 10 pGWiz.

Base pairs 3415 - 3718 are non-coding DNA that is residual from vector pNK2859.

Base pairs 3719 - 3761 are non-coding  $\lambda$  DNA that is residual from pNK2859.

Base pairs 3762 - 3831 are the 70 bp of the left insertion sequence recognized by the transposon Tn10.

Base pairs 3832-3837 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 3838 - 4527 are the multiple cloning site from pBluescriptll sk(20), corresponding to bp 924-235 of pBluescriptll sk(-). This multiple cloning site may be used to insert any coding sequence of interest into the vector.

Base pairs 4528-4532 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 4533 - 4602 are the 70 bp of the right insertion sequence recognized by the transposon Tn10.

Base pairs 4603 - 4644 are non-coding  $\lambda$  DNA that is residual from pNK2859.

Base pairs 4645 - 5488 are non-coding DNA that is residual from pNK2859.

Base pairs 5489 - 7689 are from the pBluescriptll sk(-) base vector - (Stratagene, Inc.), corresponding to bp 761-2961 of pBluescriptll sk(-).

Completing pTnMod is a pBlueScript backbone that contains a colE I origin of replication and an antibiotic resistance marker (ampicillin).

It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

All plasmid DNA was isolated by standard procedures. Briefly, *Escherichia coli* containing the plasmid was grown in 500 mL aliquots of LB broth (supplemented with an appropriate antibiotic) at 37°C overnight with shaking. Plasmid DNA was recovered from the bacteria using a Qiagen Maxi-Prep kit (Qiagen, Inc., Chatsworth, CA) according to the manufacturer's protocol. Plasmid DNA was resuspended in 500  $\mu$ L of PCR-grade water and stored at -20°C until used.

#### **EXAMPLE 2**

## 25 Transposon-Based Vector pTnMCS

5

10

15

20

30

Another transposon-based vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells. This vector was termed pTnMCS and its constituents are provided below. The sequence of the pTnMCS vector is provided in SEQ ID NO:56. The pTnMCS vector contains an avian optimized polyA sequence operably-linked to the transposase gene. The avian optimized polyA sequence contains approximately 75 nucleotides that precede the A nucleotide string.

Bp 1-130 Remainder of F1 (-) ori of pBluescriptII sk(-) (Stratagene) bp1-130

Bp 133 - 1777 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems) bp 229-1873

Bp 1783 - 2991 Transposase, from Tn10 (GenBank accession #J01829) bp 108-1316

Bp 2992 - 3344 Non coding DNA from vector pNK2859

5 Bp 3345 – 3387 Lambda DNA from pNK2859

Bp 3388 - 3457 70 bp of IS10 left from Tn10

Bp 3464 - 3670 Multiple cloning site from pBluescriptII sk(-), thru the XmaI site bp 924-718

Bp 3671 - 3715 Multiple cloning site from pBluescriptII sk(-), from the XmaI site thru the XhoI site. These base pairs are usually lost when cloning into pTnMCS bp 717-673

Bp 3716 – 4153 Multiple cloning site from pBluescriptII sk(-), from the XhoI site bp 672-235

Bp 4159 - 4228 70 bp of IS10 right from Tn10

15 Bp 4229 - 4270 Lambda DNA from pNK2859

Bp 4271 – 5114 Non-coding DNA from pNK2859

Bp 5115 - 7315 pBluescript sk (-) base vector (Stratagene, Inc.) bp 761-2961.

#### **EXAMPLE 3**

### 20 Production of Antibody in Egg White

25

30

A transposon-based vector containing a CMV promoter/cecropin prepro/antibody heavy chain/cecropin pro/Antibody light chain/conalbumin poly A (SEQ ID NO:97) was injected into the oviduct of quail and chickens. A total of 20 birds were injected (10 chickens and 10 quail) and eggs were harvested from the birds once the eggs were laid. Partially purified egg white protein (EW) was then run on a gel under both reducing and non-reducing conditions. Figure 5 is a picture of the gel. Lanes 1 & 18: molecular weight markers, Lanes 2 and 3: EW #1, non-reduced, reduced, respectively; Lanes 4 and 5: EW #2, non-reduced, reduced, respectively, Lanes 8 and 9: EW #4, non-reduced, reduced, respectively; Lanes 10 and 11: EW #5, non-reduced, reduced, respectively; Lanes 14 and 15: EW #7, non-reduced, reduced, respectively; and Lanes 16 and 17: EW #8 Control, non-reduced, reduced, respectively. Based upon the gel results, the

possibility that the egg white in the treated chicken and quail contains antibody produced by the above-mentioned transposon-based vector cannot be excluded.

#### **EXAMPLE 4**

5 Additional Transposon-Based Vectors for Administration to an Animal

The following example provides a description of various transposon-based vectors of the present invention and several constructs for insertion into the transposon-based vectors of the present invention, all for intraoviduct administration. These examples are not meant to be limiting in any way. The constructs for insertion into a transposon-based vector are provided in a cloning vector pTnMCS or pTnMod, both described above.

# pTnMOD (CMV-prepro-HCPro-Lys-CPA) (SEQ ID NO:97)

Bp 1-4090 from vector pTnMod, bp 1 - 4090

10

15

20

30

Bp 4096-5739 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864

Bp 5746-5916 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733 Bp 5923-7287 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7288-7302 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)

Bp 7309-7953 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7960-8372 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

25 Bp 8374-11973 from cloning vector pTnMod, bp 4091-7690

# pTnMCS (CHOVep-prepro-HCPro-CPA) (SEQ ID NO:98)

Bp 1-3715 from vector pTnMCS, bp 1-3715

Bp 3721-4395 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1-675

Bp 4402-5738 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336

Bp 5745-5915 Capsite/Prepro taken fron GenBank accession # X07404, bp 563-733

Bp 5922-7286 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

- Bp 7287-7298 Pro taken from GenBank accession # X07404, bp 719-730 (does not include Lysine)
- 5 Bp 7305-7949 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
  - Bp 7956-8363 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
  - Bp 8365-11964 from cloning vector pTnMCS, bp 3716-7315

10

## pTnMCS(CHOvep-prepro-HCPro-Lys-CPA) (SEQ ID NO:99)

- Bp 1-3715 from vector pTnMCS, bp 1-3715
- Bp 3721 4395 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1–675
- Bp 4402 5738 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336
  - Bp 5745 5915 Capsite/Prepro taken fron GenBank accession # X07404, bp 563–733
  - Bp 5922 7286 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
- 20 Bp 7287 7301 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)
  - Bp 7308 7952 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
  - Bp 7959 8366 Conalbumin polyA taken from GenBank accession # Y00407, bp
- 25 10651-11058
  - Bp 8368 11967 from cloning vector pTnMCS, bp 3716-7315

### pTnMCS (CMV-prepro-HCPro-CPA) (SEQ ID NO:100)

- Bp 1 3715 from vector pTnMCS, bp 1-3715
- 30 Bp 3721 5364 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864
  - Bp 5371-5541 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733

- Bp 5548 6912 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
- Bp 6913 6924 Pro taken from GenBank accession # X07404, bp 719-730 (does not Lysine)
- 5 Bp 6931 7575 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
  - Bp 7582 7989 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
  - Bp 7991 11590 from cloning vector pTnMCS, bp 3716-7315

10

# pTnMCS (CMV-prepro-HCPro-Lys-CPA) (SEQ ID NO:101)

- Bp 1 3715 from vector pTnMCS, bp 1-3715
- Bp 3721 5364 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864
- Bp 5371-5541 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733
  Bp 5548 6912 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
  - Bp 6913 6927 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)
- 20 Bp 6934 7578 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)
  - Bp 7585 7992 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
  - Bp 7994 11593 from cloning vector pTnMCS, bp 3716-7315

25

# pTnMod (CHOvep-prepro-HCPro-CPA) (SEQ ID NO:102)

- Bp 1-4090 from vector pTnMod, bp 1-4090
- Bp 4096-4770 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1-675
- 30 Bp 4777-6113 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336
  - Bp 6120-6290 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733

Bp 6297-7661 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7662-7673 Pro taken from GenBank accession # X07404, bp 719-730 (does not include Lysine)

5 Bp 7680-8324 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 8331-8738 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 8740-12339 from cloning vector pTnMod, bp 3716-7315

10

25

## pTnMod (CHOvep-prepro-HCPro-LYS-CPA) (SEQ ID NO:103)

Bp 1-4090 from vector pTnMod, bp 1-4090

Bp 4096-4770 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1-675

Bp 4777-6113 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336

Bp 6120-6290 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733 Bp 6297-7661 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

20 Bp 7662-7676 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)

Bp 7683-8327 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 8334-8741 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 8743-12342 from cloning vector pTnMod, bp 3716-7315

#### pTnMod (CMV-prepro-HCPro-CPA) (SEQ ID NO:104)

Bp 1-4090 from vector pTnMod, bp 1-4090

30 Bp 4096-5739 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864

Bp 5746-5916 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733

Bp 5923-7287 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7288-7299 Pro taken from GenBank accession # X07404, bp 719-730 (does not include Lysine)

5 Bp 7306-7950 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7557-7969 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 7971-11970 from cloning vector pTnMod, bp 3716-7315

10

20

25

30

## Example 5

Production of a Vaccine in Transgenic Chickens against Brucella Melitensis, Collection of the Vaccine from Eggs and Use of the Vaccine to Confer a Protective Response

Brucella species are considered potential biological warfare agents due to the characteristic protracted debilitating disease caused by these organisms. The antigens of interest are the Group 3 Omps which have been shown to demonstrate a role in protective immunity. Group 3 Omps are comprised of two separate proteins of 25 kDa (Omp25) and 31 kDa (Omp31) (Cloeckaert et al., 1996a).

First, genes encoding *Brucella melitensis* outer membrane proteins (Omp 25 and Omp 31) were cloned for expression into egg albumin. The vectors that are used have been described above and the genes of interest encode for Omp 25 and Omp 31.

Harvest of Brucella Genomic DNA: Genomic DNA from B. melitensis is harvested from bacteria grown in 250 ml of Brucella Broth (Difco Laboratories, Detroit, MI) for 18 hours at 37° C. The culture is killed by the addition of chloroform for 1 h to prevent accidental infection of lab personnel. After killing, genomic DNA isharvested using a Qiagen Genomic-Tip System (Qiagen, Inc., Valencia, CA) according to the manufacturer's protocol and resuspended in PCR grade water.

PCR Optimization: For PCR optimization, primers specific to Omp25 and Omp31 are used to amplify each gene from genomic DNA of *B. melitensis*. Conditions for PCR with each set of primers are optimized using a FailSafe<sup>TM</sup> PCR System kit (Epicentre Technologies, Madison, WI), the PCR fragments separated by agarose gel electrophoresis, excised from the gel, and purified from the agarose using

a QIAquick Gel Extraction Kit (Qiagen, Inc). Purified DNA corresponding to Omp25 and Omp31 is stored at -20° C until needed.

Plasmid DNA Isolation: All plasmid DNA is isolated by procedures routinely used in the laboratory of Dr. Cooper. Briefly, *Escherichia coli* containing the desired plasmid is grown in 500 ml aliquots of LB broth (supplemented with the appropriate antibiotic) at 37°C overnight with shaking. Plasmid DNA is recovered from the bacteria using a Qiagen Maxi-Prep kit (Qiagen, Inc.) according to the manufacturer's protocol. Plasmid DNA is resuspended in 500  $\mu$ L of PCR grade water and stored at 20°C until needed.

5

10

15

20

25

30

The vectors are digested with restriction enzymes. Bands corresponding to the vector are excised from the gel, purified from agarose (see above), and ligated (Fast-Link<sup>TM</sup> DNA Ligation Kit, Epicentre Technologies) overnight with either DNA encoding Omp25 or Omp31. Ligation mixtures are electroporated into  $E.\ coli$  STBL4<sup>TM</sup> (Life Technologies, Rockville, MD), plated onto LB agar plates supplemented with 200  $\mu$ g/ml ampicillin, and incubated at 37° C overnight.

Resulting colonies are grown at 37° C overnight in 2 ml of LB broth supplemented with 200 μg/ml ampicillin, plasmid DNA harvested using a QIAprep Spin Miniprep Kit (Qiagen, Inc.), and separated by agarose gel electrophoresis. Bacterial isolates containing the expected insert size are frozen at <sup>80°</sup> C. The insert is sequenced to determine orientation and verify the sequence. All sequencing is conducted using a BigDye<sup>TM</sup> Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, Foster City, CA) according to the manufacturer's protocol and loaded onto a PE Applied Biosystems 310 Genetic Analyzer. Isolates containing Omp25 (pTnOv25) and Omp31 (pTnOv31) with the correct sequence and orientation are grown as described above for Maxi-Prep DNA isolation. Purified DNA will be resuspended in TE buffer (10 mM tris, 5 mM EDTA, pH 8.0) and stored at <sup>20°</sup> C until needed.

Next transgenic chickens are developed to produce Omp25 and Omp31 in the albumin portion of their eggs. The experimental males are obtained from commercial sources at one day of age and maintained in brooders until used. After being placed in a treatment group, males will be housed individually in standard caging in temperature controlled spaces appropriate to the age of the birds. They are watered and fed standard Starter Chick chow ad lib and kept initially in a 23:1 h light/dark

cycle, stepping down at approximately weekly intervals to a 15h:8h light/dark cycle, as this regimen optimizes sexual maturity and fertility (Duncan 1999).

For transfection, plasmid DNA is complexed with Superfect<sup>TM</sup> (Qiagen, Inc) in an approximate 1:3 ratio, following the manufacturer's protocol. The DNA is injected (described below) at a rate of 5  $\mu$ g per testis in a 50  $\mu$ L volume [Hank's balanced salt solution (HBSS) will be used to bring the total volume to 50  $\mu$ l]. At least 20 males are used for controls and receive an injection of HBSS and transfecting reagent. All birds are held until sexual maturity before sperm samples are obtained for PCR.

5

10

15

20

25

30

At age 30-98 days, groups of individual males are subjected to transgene delivery by direct intratesticular injection of DNA. Males are anesthetized with isoflourane via a gas machine and positioned on their side on an animal board with cords tractioning wings and feet to allow access to the testes area. The area is swabbed with 0.5% chlorhexidine and a 2 cm dorsolateral incision made over the testis (which is identical to the procedure for caponization). After exposing the testis, a capillary injection device (described below) is used to inject, in 2 or 3 sites, about 50  $\mu$ l of the DNA/SuperFect complex (Qiagen) (~5  $\mu$ g DNA per testis). The injections are calculated to suffuse the DNA throughout the whole testis, allowing as much contact of DNA with spermatogonia as possible. Following injection, the wound is closed in two layers with 4-6 absorbable suture, and the contralateral testis similarly exposed and injected. Following surgery, the bird is returned to its cage to recover. Approximately 50 males are used for this experimental regimen to increase the likelihood of obtaining a male having transgenic sperm.

The injection device consists of a 27 ga hypodermic needle cut off at the hub, bent to the desired angle, and welded onto a standard 100 µl capillary tube. The capillary is filled with the DNA solution and placed into a standard mouth pipette so that the injection rate and volume can be accurately controlled hands-free. After injection, males are allowed to reach sexual maturity in order to obtain sperm for PCR. Males with sperm positive for Omp25 or Omp31 are mated with females to produce transgenic F1s. The F1s are verified for copy number of the transgene. These transgenic F1s are raised to maturity.

Next the expressed proteins are purified from the eggs of the transgenic chickens and subsequently used in vaccine efficacy trials. The purification

is achieved with methods presented above for purification of proteins from eggs. The following is only one approach to purify the protein. The initial objective in the purification of expressed proteins in egg white is to eliminate the major component of egg white, namely, the albumin proteins. Li — Chan, et al., 1986 have described a versatile and simple ion — exchange procedure which largely accomplishes this elimination of the albumin.

5

10

15

20

25

30

The egg white is lightly homogenized to reduce viscosity [shear rate of 3140 sec<sup>-1</sup> reduces albumin viscosity from 21 to 5 centipoises (Tung et al., 1969]. The albumin solution is filtered through two layers of muslin to remove chalazae. The resulting solution is then passed through an industrial cation – exchanger, IMAC-HP333 (Rohm and Haas, Philadelphia, PA) which has been equilibrated at pH 9 with low ionic strength buffer (phosphate or bicarbonate). The albumin passes through the ion – exchange column leaving some ovomucin and all of the avidin and lysozyme on the column.

This procedure allows cationic proteins/peptides to be separated from the albumin at an early stage in the purification. What is required for making this process a general one is to either purify such proteins which are by their nature, highly positively charged (e.g. cecropin – like lytic peptides or their derivatives) or utilize fusion proteins wherein the fusion adduct contains the requisite cationic charge. Also required is that the magnitude of charge on the expressed protein or fusion protein must exceed considerably that of any possible cationic contaminant.

Once the cationic entities have been captured by the ion – exchanger, elution is initiated by a series of increasing ionic strength, pH 8.0 bicarbonate buffers. The desired product is retained until it is selectively eluted at the highest ionic strength (ammonium bicarbonate, approaching 2M). The water and most of the ammonium bicarbonate are removed by lyophilization. The recovered fusion protein is dissolved in 88% formic acid and subjected to cyanogen bromide cleavage. If the desired component is anionic, the lyophilized residue is dissolved in dilute acetic acid and passed through and anion – exchange resin contained in a jacketed column. The column is washed with dilute hydrochloric acid (0.05 M) and cold (-5°C) ethylene glycol solution pumped through the jacket on the column. By-products from the cleavage reaction are substantially removed by this step. The anionic protein is then lyophilized (using solid KOH in the vacuum system to capture the HCl) to dryness,

5

10

25

30

dissolved in a minimal amount of water and lyophilized (twice) to remove residual HCl.

If the desired component of the fusion protein is cationic then the cleavage mixture is diluted 5 – fold with acetate buffer and passed onto a sulfoethylcellulose cation exchanger contained in a jacketed column. The column is washed with 0.05 M HCl and then 0.15 M HCl to remove residual cationic materials including degradation products while pumping cold (-5°C) ethylene glycol through the column jacket. The desired product is eluted with 1.0 M HCL adjusted carefully to pH 2.7 by the extremely slow addition of very cold ammonium hydroxide. The eluted protein is lyophilized to dryness using solid KOH within the vacuum system to capture the HCl. The protein is dissolved in a minimum amount of water and lyophilized (twice) to remove residual HCl. Protein quantity will be determined using a Bio-Rad Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA). Purity is determined with Western blots, HPLC, ELISAs and other methods known to one of ordinary skill in the art.

Vaccination and Challenge of Mice: The BALB/c murine model of brucellosis established by Montaraz and Winter (1986) has provided much of the current understanding of immunity to brucellosis. The model is valuable since the murine colonization profile of virulent type strains and vaccine candidates mimics what is often attenuated in the ruminant host (Montaraz and Winter, 1986; Winter, 1990; Schurig et al., 1991; Elzer et al., 1996; Stevens et al., 1997; Edmonds et al., 2000a; Edmonds et al., 2000b). The murine model is also useful due to practical concerns including availability of reagents, cost, and population homogeneity (Montaraz and Winter, 1986; Garcia-Carrillo, 1990).

Bacterial Strains: Brucella melitensis strain 16M has been documented to cause chronic infections in mice (Jiminez et al., 1994; Philips et al., 1995). Infectious doses containing approximately 1 x 10<sup>9</sup> and 1 x 10<sup>11</sup> CFU/ml are prepared as described (Elzer et al., 1994c) and snap frozen in liquid nitrogen prior to storage at -80°C. Immediately prior to animal inoculation, samples are thawed and diluted in sterile saline to the appropriate concentration. Doses are verified by serial dilution and plating on Schaedler agar containing 5% bovine blood (SBA) (Difco Laboratories, Detroit, MI).

Female BALB/c mice at approximately 6 to 8 weeks of age are obtained from the Division of Laboratory Animal Medicine at the School of Veterinary Medicine at Louisiana State University (Baton Rouge, LA) and held at least one week prior to use. All animals are housed in a restricted access facility. All mice are used in accordance with the Animal Care and Use Committee at Louisiana State University, and approved protocols will be followed.

5

10

15

20

25

30

Murine Vaccine Efficacy Trials: Mice are divided into groups of ten and are given the following vaccines intraperitoneal (i.p.): Group 1 – saline control; Group 2 –  $5 \times 10^4$  CFU of B. melitensis vaccine strain Rev-1 i.v.; Group 3 – Titremax adjuvant; Group 4 – Omp25 plus Titremax adjuvant; Group 5 – Omp31 plus Titremax adjuvant; and Group 6 – Omp25 and Omp31 plus Titremax adjuvant. Mice will be vaccinated with  $30 \mu g$  B. melitensis Omp25 or Omp31. The dose of the Rev-1 strain is based on previously published work in BALB/c mice (Jimenez de Bagues et al., 1994a; Jimenez de Bagues et al., 1994b). Prior to vaccination and 8 weeks post-vaccination, mice were bled and the serum pooled for each group and the Omp25 and Omp31 vaccinated mice are boosted.

Ten weeks following initial vaccination with either Omp25, Omp31, Omp25 plus Omp31, or *B. melitensis* strain Rev. 1, both vaccinated and non-vaccinated mice are i.v. inoculated with the challenge strain 5 x 10<sup>4</sup> CFU of *B. melitensis* strain 16M (Jimenez de Bagues et al., 1994a; Jimenez de Bagues et al., 1994b). Two weeks post-challenge, all mice are euthanized by halothane overdose. At necropsy, the whole spleen is removed by aseptic technique for bacteriological analysis while blood is obtained for serological analysis. Spleens are homogenized with a Sorvall Omni-Mixer (Newton, CT) in sterile phosphate-buffered-saline (PBS) and subsequently serially diluted and plated on SBA. All plates are incubated at 37°C in a 5% CO<sub>2</sub> atmosphere for 3 to 4 days until isolated colonies appear.

Serological procedures: Blood is collected from all the mice before inoculation with *B. melitensis* and at two-week intervals thereafter until they are killed. The sera obtained is stored at 80°C. *Brucella*-specific antibodies are detected by the Card test and Western blot analysis with whole-cell lysates of *B. melitensis* (Elzer et al., 1994a).

PCT/US2003/041269 **WO 2004/067743** 

Vaccination and challenge of Goats: Strains

5

15

20

B. melitensis is of established virulence in pregnant goats (Meador and Deyoe 1986, Enright 1990). The construction and phenotypic characteristics of the isogenic 2308 htrA mutant, PHE1, have been described by Elzer et al., (1994a). Infectious doses containing approximately 109 or 1011 colony forming units (cfu) ml-1 in Brucella broth (Difco, Detroit, MI, USA) were prepared as described by Elzer et al., (1994a) and snap frozen in a dry ice-ethanol bath before being stored at -80°C. These preparations were thawed just before use and their content of viable brucellae was assessed by serial dilution and plating.

Serological procedures: Blood is collected from all the goats before they are 10 vaccinated, and at two-week intervals thereafter until they are killed. obtained is stored at -80°C. Brucella-specific antibodies are detected by the Card test and Western blot analysis with whole-cell lysates of B abortus (Elzer et al., 1994a).

Vaccine Efficacy Study: Seventy-two mature, non-pregnant female goats are divided into 6 groups and given 100  $\mu$ g of the following vaccines subcutaneously (s.c.): Group 1 - saline control; Group 2 - 5 x 10<sup>4</sup> CFU of B. melitensis vaccine strain Rev-1 i.v.(this group receives saline first and then Rev-1 2 weeks later); Group 3 - Titremax adjuvant; Group 4 - Omp25 rlus Titremax adjuvant; Group 5 - Omp31 plus Titremax adjuvant; and Group 6 - Omp25 and Omp31 plus Titremax adjuvant. Eight weeks later, the 3 Omp groups and adjuvant control group are boosted with a second dose of the appropriate formulation. Twelve other animals are vaccinated SC with 1 x 106 CFU of the vaccine B. melitensis strain Rev. 1. Previous experiments in goats have shown that this dose of strain Rev. 1 provides protection against challenge with virulent B. melitensis strain 16M (Elzer et al., 1998). The remaining 12 animals serve 25 as controls and receive another SC inoculation of the saline diluent. Four weeks after vaccination, a fertile male goat is housed with each group. Following ultrasound confirmation of pregnancy, all animals are conjunctively challenged at approximately 110 days gestation with  $1 \times 10^7$  CFU of virulent B. melitensis strain 16M. Following inoculation, animals are monitored daily for delivery. At parturition, the birth status 30 of the kid is recorded and within 24 h, necropsy samples obtained for bacteriologic analysis from the lung and abomasal fluid. Within one month of delivery, dams are

euthanized and samples taken for bacteriologic analysis from the parotid lymph node, prescapular lymph node, supramammary lymph node, internal iliac lymph node, liver, spleen, milk, and uterine fluid. Tissue samples are frozen at -4°C for later analysis; milk and uterine fluid are swabbed the day of collection. Serum samples are obtained prior to vaccination, at challenge, and at necropsy for serological analysis.

5

10

25

30

Bacteriological procedures: All the tissue specimens are homogenized in sterile PBS (0.002M NaH<sub>2</sub>PO<sub>4</sub>, 0.008M Na<sub>2</sub>HPO<sub>4</sub>, 0.15M NaCl, pH 7.2), plated on to Farrell's selective medium (Farrell 1974) and Schaedler agar supplemented with 5% sheep blood. The plates are incubated for 14 days at 37°C in an atmosphere of 5% carbon dioxide. The isolates were identified as *Brucella* species by their Gram-staining characteristics, colonial morphology, urease and oxidase reactions and agglutination with *B. abortus*-specific antibodies in a standard slide agglutination assay. (Elzer et al., 1994a)

Mice Studies: Mice vaccinated with the subcellular components (OMP) in conjunction with the adjuvant Titermax exhibit increased protective immunity (i.e., lower CFU/gm of tissue) as compared to the saline controls. Serological analysis remains negative since the *Brucella* diagnostic tests are dependent on antibody reactions to the O-side chain of the organism. The adjuvant alone does not contribute to the immune response or interfere with the serological analyses. The combination of the Titermax and OMP vaccines results in a predominant TH1 protective immune response as demonstrated by a lower bacterial load in cultured tissues.

Goat Studies: Goats receiving the Titermax and OMP combinations exhibit a protective immune response post-challenge as demonstrated by decreased colonization of the dam/fetus pairs and/or decreased abortions as compared to the saline or Titermax alone controls. *B. melitensis* Rev 1 is the accepted vaccine for caprine brucellosis; it is the "gold standard" for protection in this model. The Titermax/OMP vaccinations results in equivalent or better protection as compared to Rev 1. However, unlike Rev 1, the Titermax/OMP groups remain serologically negative on all diagnostic tests until after challenge when all groups become positive.

All patents, publications and abstracts cited above are incorporated herein by reference in their entirety. It should be understood that the foregoing relates only to preferred embodiments of the present invention and that numerous modifications or alterations may be made therein without departing from the spirit and the scope of the present invention as defined in the following claims.

Appendix

SEQ ID NO:2(cecropin pro extended) GCG CCA GAG CCG AAA TGG AAA GTC TTC AAG ID NO:1 (cecropin pro) CCA GAG CCG AAA ggg 10

S

CTG GCT TTG TCA ACA GII SEQ ID NO:3 (cecropin prepro)
AAT TTC TCA AGG ATA TTT TTC TTC GTG TTC GCT TTG
GTT TCG GCT GCG CCA GAG CCG AAA

ID NO:4 (cecropin prepro extended) TTC TCA AGG ATA TTT TTC TTC GTG TTC GCT TTG GTT CTG GCT TTG TCA ACA TCG GCT GCG CCA GAG CCG AAA TGG AAA GTC TTC AAG SEQ AAT GTT 15

SEQ ID NO:5 (modified Kozak sequence) 20

ACCATG

atg ctg ggc atc tgg acc ctc cta cct ctg gtt ctt acg tct gtt gct aga tta SEQ ID NO:6 25

atg att cct gcc aga ttt gcc ggg gtg ctg ctt gct ctg gcc ctc att ttg cca ggg acc ctt tgt SEQ ID NO:7

gcagagcaat ggtggccagg ctgggggctgg ggctgctgct gctggcactg ctcctaccca cgcagattta ttcc SEQ ID NO:8 atgg 30

atgaatctat cgaacatttc tgcggtaaaa qtattaacac tggtggttag cgctgccatc gct SEQ ID NO:9 35

cetgagaact teaggetect gggeaacgtg etggtetgtg tgetggeeca teactitgge aaagaattea ecceaecagt geaggetgee tateagaaag tggtggetgg tgtggetaat geectggeec acaagtatea etaaagetege tttettgetg tecaattet attaaaggtt

	OF ON OF CAR				
	atgaccatco ttttccttac tatggttatc		tcatacttca aqcttqqctt	gttgcatgaa acccaggtct	agctgccccg tcggacccac
ν.	acyanayany congresons addactetta aaageetaae	tgggcccaat			atcactggcg
			ctagatgaag	atcaggacat	ccagcccagt
	gaggaaaaca aggatgcgga	cttgtacaca	tcccgagtca	tgctgagcag	tcaagtgcct
			gagtacaaaa actacctgga	actacctgga	tgctgcaaac
•	atgtccatga gagtccggcg	tcactctgac			
2	11. OM 41 000				
	SECTION IN THE CONTRACTOR TO ACCOUNTAGE CONTRACTOR CONT	tctcgctttc	gctcttttct	cgtcggcgat	tgctttctcc
	gacgacgate cgttgatecg		acaagttgta tcgggaaacg atgacaacca tatgttaaac	atgacaacca	tatgttaaac
	gccgagcatc acttttcact	ttt			
15					
	SEQ ID NO:12			;	
	atg tecatettgt tttatgtgat atttettgea tatettegtg geatteagte aactaaratg	tgat atttctt	gca tatotto	gtg gcattca	igte aactaatatg
	gatcaaagga gtttgccaga agattcaatg aattctctca ttattaaact	agattcaatg	aattctctca	ttattaaact	cattcgggca
20	gacatcttga aaaacaagct	ttctaagcag	ttctaagcag gtgatggatg tcaaggaaaa	tcaaggaaaa	
		ccaccaggag	atggatggag	atgaaaatgt	gaaatcagac
	ttccagccag ttatttcaat	ggatacagac	ctcctaaggc	agcagagacg	ctacaactct
				caccactgta	cctcacagag
Ċ	gattatgttg gaagttcagt	ggtattaaac	agaacctctc	gaaggaaaag	gt
<b>C7</b>	SEC TD NO.13				
	atoutocate toactectoa	a ggagaagtct	ggagaagtet geegttaetg eeetgtgggg caaggtgaae	ccctgtgggg	caaggtgaac
	atagatgaag ttggtggtga ggccctggggc	gecetggge	aggctgctgg	tggtctaccc	ttggacccag
	addttctttd adtcctttgg	ggatctgtcc		ctgttatggg	caaccctaag
30				atggcctggc	tcacctggac
		cacactgagt		gagotgcact gtgacaagot	gcacgtggat
		242244		otastatata tastagasaa taactttaga	tcactttqqc

cetttgttee etaagteeaa etaetaaaet gggggatatt atgaagggee ttgageatet ggattetgee taataaaaaa eatttatttt eattge

SEQ ID NO:14

S

gcatggggac ggcgcttctc cagcgcgggg gctgctttct cctgtgcctt tcgctgctgc tcctgggctg ctgggcggag ctgggcagcggg

SEQ ID NO:15

10

cgaaacgatt caaaacctct ttactgccgt tatttgctgg atttttattg ctgttttatt tggttctggc aggac

SEQ ID NO:16

acaccatgga aaatgagcag gtgaatccaa agaagagct gcaaaagtta gcgcggagag daactetect aagactccgg tactatccag ctgtacctgc ctgtgcaaga cacacgatga gcatccagga ccagcagggg agagtccctg ttgggtccgc tggtagcacc caagaagacc ggagtctggg ggaggcttag tgcagcctgg ctcctgcaga atgacatgtc atagtgatgg gagacaatac ccttgtatta ttcccttccc gcagccatta atcatctcca aaactaaaac gaggacacag acacagagtt gtgaaat tgaatacgaa gagacgattc tctgaggtct ctgtgagctc ggagttggtc

SEQ ID NO:17

20

caagccagrc tgcattgctc ccctgggagg tttaaaatat tcactgcgga actacaccca agaagtgtgc ggaactggcc caacagtgtg gcagcaaacc ttcactcgaa gagcagcgtg cagacctctg ggtctcagag atcataataa aatcagaaat cccagggtac cgccaaagtg tgcatgactg tggcaaccag acatccattc ttctgtgagt agctcaccat ggctggacct cttcacggac acaggctgcc cctcttgctt gtcacatctt aagcatatt ccagacttgg

25

SEQ ID NO:18

30 tggaagcaag agggagtatg ctaacttcat

ρ

SEQ ID NO:19

atcaattaca agaggg

cccccggat ccatggccgc taaattcgtc gtggttctgg ccgcttgcgt cgccctgagc cactcggcta tggtgcgccg caagaagaac ggctac atgaagttmg catactccct cttgcttcca ttggcaggag tcagtgcttc agtkatcaat caaaccggac tttgcttatc cttttagtga agatgatacc cccccggat ccatgaaact cctggtcgtg ttcgccatgt gcgtgcccgc tgccagcgct ggatcaggat agaagtctga cgtttacggt aagetteteg tgaaaaceaa eecaattagt tagtattgea ttetgtgtae tatagtttgg ttacaaaatc catgttggca atttgttttc gtcggtaccg agcttcacga tgcgattttg ttcacaggca ggtttttgta gagaggggca tgtcatagtc ctcactgtgg aaccatatg aagtgtcacc ttccaagcat tgtttgcacg tgtagcattt aattottaat taattattgt ggtgtcacaa taacttttc cagtgtacgg cggctcgagg cagaagtccg gacgcata gaattcatta tcagcacgga ccagatttct tegttgegtt aatattaaaa atattttaaa atacctccat tgcaaaagac atggctaaag ttatgattgt ttcagttcgt gtetteteag tegttagegt tttcgaagca cagaacgcat cactctcgat tcgtagagcc ggtattccag agtgcaaccg ccatggatcc cccttccgct SEQ ID NO:21 SEQ ID NO:26 SEQ ID NO:22 SEQ ID NO:23 SEQ ID NO:24 SEQ ID NO:25 SEQ ID NO:27 tacaagaga S 10 15 20 25

30

taatgttgac

attggattt

ttgccttgga

ggatgggaaa tctgttaagt aagtactgtt

cgaagtgggg agctaatggg

tttatcattt

ggaagat

aagtggccct ctctgtttct

cttcttccca

SEQ ID NO:20

atggctacag gctcccggac gtccctgctc ctggcttttg gcctgctctg cctgccctgg cttcaagagg gcagtgcc SEQ ID NO:28

S

atgaggtct ttgctaatct tggtgctttg cttectgeee etggetgete tgggg SEQ ID NO:29

10

atgc acctgagaat ccacgcgaga cggaaccctc ctcgccggcc ggcctggacg cttgggatct ggtccctttt ctggggatgt atcgtcagct ct SEQ ID NO:30

atg gccattagtg gagtccctgt gctaggattt ttcatcatag ctgtgctgat gagcgctcag gaatcatggg ctatcaaaga agaacatgtg atcatccagg ccgagttcta tctgaatcct SEQ ID NO:31

gaccaatcag gcgagtttat gtttgac SEQ ID NO:32 15 20

aggggggatc cccggagacc ttcggggtagc aactgtcacc ttgatgctgg cgatcctgag atg gtgtgtctga ggctccctgg aggctcctgc atggcagttc tgacagtgac actgatggtg ctgagctccc cactggcttt ggct ctcctcactg gctgagggc SEQ ID NO:33 25

tgttgagcga attgttagaa catcttctgg agaaggacat tctgtgaagc ggdagtttca gccagaatgt ctcaccaccc tttctgggct tgagctgtga ctacgacttg agictccctt ttccggtgtg gatattggga gtgtgtggca tgcatcctca tcatcaggaa actctaaaaa agaaccgagt 1 gaacgattta aggagcgaat actactggta aactaatgga agaaatctgc tgcaccactg tttgatgcct catcaccttg gaaatgaggg agctcatcca ggccaaagtg tccccaagct ggatatgttg ggaactcctc aacttgctgc ctaagagggg gccacctgga ggtgctagcc aaacagctgt actgagggag accaagcaag tcagcatgta ctcccaccgt SEQ ID NO:34 121 181 241 301 30

gatgtctgag gagcatcaga gtetttttee gggtgcgagg tcagacatga cgaggttcct cacgtggccg cctggcacag cacctctacc aagtggaagc tttctgttct tetetgeett agagccagcc gtg:tcagtt tagagtgaga aagcaaacat ttcagtactg caagtgaagt cgcagcctgt tctcaaagat tcaacctgca atgttattgt cactggtgtt gggatgtgga ctgcacaccg ccatctatgg gaggtgctat ctctatgaga aacactccct aattttatca ttctatgtga acgccaactg ctgatcgtgg ggttgaccaa caagatggaa agaaccacgc aggatccct gcccacgcgc agtgccatac gaacaccaaa ttgtgatatg tetgtgeege gtgcccatca ctttgcctgt aagagcctga cattagttaa gtgagttaac ctcaaggcac ttgttaatga aggatcggga aggttatgct cccatattt ctttttgtcc ggctctgaag cccatggaag gcctcgtggg ccctgcttca catccacgcc ggatgcacgg tttcattcaa atctgtctag ccacctctct gcacagttac ctctttgaca gcctctcttg cgtggcctag cgctctggag gacgtccatg gtggaggcg gactgaagtg qatactqtqq ctgagcgggc actgcagcac gctgcagttg atgtagttct ccgccargcd acctccccat ttgatcttca gcttttatta actgaaatga gactctgtaa ctattcccac tagcttatct ctgccaccaa agatgagact cttgtttcct ctgtgaagcc catataaatt atceteegee attatcttt gaagcggcct ctcgtttggg ttgatctttg cttttatttc ggtttttcag tgctgccacc actggaattt tttgggctat gcagaatttt ctcgcatggt cctgtcgaca gtctcggcct cttcatccag ggtgaagcct gcacttatca cggtaaagaa aagttgccga caagtgtttt atgtctgaat acatgatgtc tttagtgatt gagctttgtt tgtttcccta aaaagcttcc aaagggactg gatagageet agctcctttt gaatgtttca tttcattcat taatgtagac gcagtgtcca acactcttgg ttgtacctta ctctccctta caaagatgta tattatgtgg atctcccaga gggagagagg ctgttggctg gggcacetec ttetgtteae agctccaaga gagagaaaga caaaaatgtt tetteaaget aagagaaact tggcactcct ccctttaca agaagctccg tetgeettea ataggttgca tgcctgcctc tctcctttga ggctcttgct taaggtgaac gtgcaaggag acaccctccc ggatgtggga tagatggcac ctatctttta tgtccaaatt agatttctag tgtaaatgct caggaggtgt actccaggcc ctattgccag gcaccttgct cagtggaagg tgccttgcat ttgtttctct acgtccatct tcatttcatc gtgtggttat aagaatggtg caggaaatgc aaactgctcc cagaacaaac cagctggcat cacttcactg ctagtcaccc tectgeateg gatggtcctg tatgcggcta agagtgatgc tacaggatca agtattaaac ggtacatcga tgttcccagg cactggacca gccccctcag tgagtgtggg ttggttggac agaccatctc gactagagad cagcgccacc cttcgttgg ttcctgagaa acatgctggt aattccaccg tgtaaacaca ccagcttttg gtttggaagg gaaggccagg cttcatatct ttctttacct cagccaacat agtcacggac tgtggatggg tgagtcctgt gagcaatgtg ccagactgca cccaagccta tggatccctt aacacacttc ccacagtact agacaataaa 1021 2401 2461 1081 1141 1201 1261 1321 1381 1441 1501 1561 1621 1681 1741 1801 1861 1921 1981 2041 2101 2221 2281 2341 2161 601 721 841 481 541 199 901 961 781 5 10 35 15 20 30 25

ccggctaatt cgtgagccac gcccaggctg acgccattct cttgatctcc tatttattaa gttgtttgct gaactcctgg atgagccacc cttagaaccc tatgagaact tecagecega ggccttctca catttctgtc tcttgacaat tcttgtcttt catcatgtac agggaagaag ggcattttta acaatataca ggtcuctage tggtttgtga aatgctcata cccaacttcc tecegggtte tegetetgte gccaccgtgc ccgggatggt ggattacagg cttgctaagc aaacttccat atttttttg gactggtctc gcttacaggc caccccctg caccttggtt ccttttcctg tttacttgtt cttcctcaag ttttattcac tgaataaag gttccttttg tttgtttggg tgtcaccaca catctcccat gaagtatttg tttatccact gtgccaaata agacagagtc tctgaaactt aagctccaac acaggtgccc accatgttag caaagtgttg aaaaaaaa ttggttttta tcagcctctc aaatagctgg taaatctgat gtaaactctt atggtgattt tetteetge agttctcact atgtttctca tggtgtcctt cttctgtgtc cagtttatat atatgctact gagggcagaa accttgtttg aaatgtttgt ttgataaat gtgttttgct acttccttca tcaataaata acatagtatg ttgccatcgg tecetgetta ttttttttg agctcactgc acagggtttc ctcagcctcc cagagtagtt tgtttaaata agctgggact ttttaaaaa catttaaaat aattttgtgc acctcattaa ttgactcagg ggccttcatt gagcagtcta atcctacctt gtgtgagata gtcaaggtgc ctttgggtcg gtttttccat ggattttgta ttttaaggtg tttagtagag tctcacataa gaaaccttta tttttttt gcacgatctc atccacccac caagaccttt tettecegee ccttccaagt aggatttggt tattaccct ttccccttga ctcagcttct cccaccactc ttcagtccct gcatttattg ggctgtaaat cttagctaca tttgtacttt gcacatagtg gacatatcat caagtcctgc gcctgtagta ccactgcca aaaaaaaa tttcaaccta tttttgtatt tgacctcatg tgcgcccggg agactttttt gagggcagtg cctgcctcag gttttttgtt cctcaagcca acacctggcc ttetgettte ggttcttgcc gtggtctcct gctctccacc tecegggeee tccagtctgt teceteatt cggattatta caagtgcttg tcccctacca gcaagccaac gctccttgaa aaaatgctag gccttgctgc ccttccact tttattaaaa 2581 2641 2761 2821 2881 3061 2701 2941 3001 3181 3121 3241 3301 3361 3481 3541 3421 3601 3661 3721 781 3841 3901 3961 5 10 2 20 25

SEQ ID NO:35

gagcaccaga aagcccagca gagaaaggaa tccaagaagc caccagctaa actgcagcca ggcagaggag ggaagaggtc ttcctcctaa ctactcctca gcatgctctg gatggacatg gccatggcag gttccagctt cttgagccca tccagatcat ctgtcctcac caccaaggcc atggtgtctt cagcgactat ctgcagtttg tcaggttcaa tgctcccttc gatgttggca tcaagctgtc cgagctctgg aaggctggct ccacccagag gacagaggac aagcagaaga tecetgtact taccagcagc atggccgggc cctgggaaag tttcttcagg átatcctctg cagctaacaa gtaaccactg acaggactgg gagctggaaa aaagaggcgc 121 181 241 301 361 30 35

gcaagaactc acatccagct tctgcctcct ctgcaactcc cagcactctc ctgctgactt

atgaagetge ttgcaatggt tgcactgetg gtcaccatet gtagectaga aggagetttg 481 acaaataaat gttcaagctg SEQ ID NO:36 gttcggagac S

atgaagctgc ttgcaatggt tgcactgctg gtcaccatct gtagcctaga aggagctttg SEQ ID NO:37 gttcggagac 10

atggeett gecaaegget egaeeeetgt tggggteetg tgggaeeeee geeeteggea geetectgtt cetgetette ageeteggat gggtgeagee e SEQ ID NO:38 15

atgeegegee tgtteteeta cetectaggt gtetggetge teetgageea actteeeaga SEQ ID NO:39

atgacage atcacttgte gttttaceat egetttggtt aatattaatt atttttactg caccctatac tcactgt gaaatcccag gc SEQ ID NO:40

ttgtcagagg gactgtctca catgccagga gaagctccac ccagccctgg acagcttcga cagagcatgt gaatgccccg aggctggtga atgaaagte etgetttgtg acetgetget geteagtete ttetecagtg tgtteageag tgcatcctcg agtgcgaaga gaaggtcttc cccagccccc gtcatggcca ggagctcttg gcagctcagc cctgccgccc ctctaccagc cgagagcttc ggagatgcag catctgcggc agtecggage ttgttecagg ageaggaaga geeegageet ggeatggagg gatggagcag aagcagctgc agaagagat SEQ ID NO:41 atgcaccaag ggcggctgct cctggaggtg 30

20

SEQ ID NO:42 aattcatgaa gtgggttact ttcatctct tgttgttctt gttctcttct gcttactcta gaggtgtttt cagacg

SEQ ID NO:43

S

atgaagtggg ttactttcat ctctttgttg ttcttgttct cttctgctta ctctagaggt gttttcagac gc

10 SEQ ID NO:44

tcattttcag actttgtact actccctgct gctggtgtta tttagaatac acaaaaa.ca gagaaagaaa acacactctg gatagageet tctcccatct ctgcttttqt ggatcccact tecetetgte atacacac acacacac acacacaca acacacaca cccaggtacc gcagcctagt agaagaataa tcacccgagt agtttgctgg ttagcagacg ctcatcgcag ccctccata gctctagaat gatecttatt ttetatggea aagggaaaaa gttatttgtc accatgccca attttgtgta caaatgttcc tggcaaaggc aagtgtacat tataaatagc aaaacagctg gcttggacca ctgtttgcta aatctcactg tcactgctaa attcagagca tatttaagtg tggggaattc gctctcaaca aacaggaaaa ttgtttgccc tagtetttge gtaaggtgta atgetgeetg aaaattgtag ttgttttaaa taatgcattt agagagaagg ttttgtagat aaatgtgagg ggacaagggg acattettt tgttgccggc cagtcaccca gttgagggat ttgaatgaca tcataaccct caaaattgaa atgtgacatt gctaaccaat ccaacccaat acagtgacag ttgaagtcat aacgtttcta agggaggact cgagacacgc cagctcggca gaccaatcca ttggcaactg cttcagggtt tttaatactg tcagaaaca tcagtaacac aaggtatgat agagggagag cattattcct ttcttactct ttccgtagta aaattcaaac cagcagtett aaaggggttc tgttggatgc tggggaggg ttgggtttta cacccctcac cccatattcc ttaaaatgca tactcaataa ctttgccaga tectgtetae agtgtetgtg gccctaaagg acccagtca agaaagttaa cttctttt gcgaaccctg ataaagtcct gttatatgga agtggtgggg ctttttgctt tgggaaaaat aggtaaatat catctttcat acccataaag aaacacaggt tctatttcag atgagacagt aatgagatca aatctacttt tctcccatcc atcgtccttt catgcccagc tgctctagtt tttaatgtct aatccctctt gcgcaatgga ctctggattt tcagaagcaa gatttcttga gtaataatag ccaggggaat ttgctagcca gcacacagac agaggtttga cttctcccag cgttctagga agaaaaatgt gagaaaagt gctgtcttcc ctaaaattgc gcataagtat agagggtctt tttgtgtctt aatgcttctg cccagctgtt aaaagtcct gaatteteaa 121 1261 1081 1141 201 1321 241 301 1021 1441 181 361 721 421 661 601 781 841 901 481 541 196 15 20 25 30 35

atgaa gtgggtaacc tttatttccc ttcttttct ctttagctcg gcttattcca ggggtgtgtt tcaaggagga ctcctcgcat acagtcgtgg eggtaceege tgeagggete caaacaggae atgacteget eegeettgge egagetgete ggagaacgag gctctggagg aggagaactt ccctctggcc ggggcctctg ctcgccccc gggagagaaa ggccggctgc aagaacttct tctggaaaac cttcacctcc gectecatea getgeteett egeeggaeag agagaeteea aacteegeet getgetgeae 1 atgaagatgg teteeteete gegeeteege tgeeteeteg tgeteetget gteeetgaee ctccagtcaa cactacaaca gaagatgaaa cggcacaaat ccggctgaa gctgtcatcg gatttcgatg ttgctgtttt gccattttcc aacagcacaa ataacgggtt attgtttata aatactacta ttgccagcat tgctgctaaa gaagaagggg aggctgctcc agccagcact ggtcctatgg actgcgccct ggaggaaaga gagatgccga atgagatt toottcaatt tttactgcag ttttattcgc agcatcctcc gcattagctg atgaag ccaattcaaa aactcctagc tggccttatt ctactgactt cgtgcgtgga ctgattaatg cttgcaagga acaccatgtc ctgccacggc gaaggaaggac ccgaggacgc ccacgccgac ctagagcggg ccgccagcgg gggcactggc ctgggtagtc gtgaagatgc cccactccc ttcaccagct gtactctatt ggcttctgtg tggcgctgtg cctgctcacc ttctgtccca tcttgggcag tatctttgga taaaagagag gctgaagctt tccctctgtt gtcaatgcac ctgtcggacc tcctgcaggg gttacttaga tttagaaggg ctcttctacc gggttctcat aatcgtggga aacctgaacc SEQ ID NO:45 SEQ ID NO:46 SEQ ID NO:48 SEQ ID NO:49 SEQ ID NO:47 tegtegaga aaattt 1621 1681 1561 121 181 241 361 301 61 25 10 15 30 5 20

gaggcggcgc

cgagctgcgc

ccaggagccc gggggcgcccg cggcgggcat cggcatctcc ttcgagtacc accgctaccc

gctctgtgcg

atggccgggc gagggggcag cgcgctgctg

gcgccgaagc

agcaagagga

aggaggatga tggctcctgg

cacggtgggg ggctgttgga ggggaacgag gaacccagat gggtcgaage agtgtacgtg atatgatgag tttccaaagc tgcttggtac ggatattcct gccaccatgt tgagatcacc aattgtggat aggatttgtc ctgggaggat aaactataaa agaggagaag gcaggattta ccgacaaccc tggataggat atatgaagaa atgacgccat aataccagaa tgccttccct actggtttgt attggattat ccaattatcc caacggtgg acccaatcg gcaactgttt tgaagaccta ccgtggaagg gaggagttaa ttatacctgg aagtggcagt aaaggaaaga atcgagctgt gggaatatgc ctatgcaacg attcacatca gaactcaagg ctgttgaaaa gctgtcattc gaccttgtgg tecteceag gccatgtctg accgccatca tttccagacc gatggaacca taccttagca gaagagactc cagatacacc gccaccatct tggagattgc ataacaaaga tcagaaactt gctgcagtgc gctcctggtc ggcccagtac taaatacatt cagtacccgc tcagcctggt gaaccggaac cgaatacagc aaataatcat tgagaccaag ccatggagga tttcaacccg cagctttgta gttcccacct ctaccttgag aattgcgaat tggtgattac ctatctggca tgaactggag agacttcaat gaaaatgatg tgtccgtgtg agggccggga agcctgaatt tcattttctt aggcagcgtc gaatagatct aaggtggtcc agcttgctcc gtagtgctca acctgatcca catactette atgatgacag ctgccaatct gagggatgca gctgtgagaa ccctcattag aaggtaaccc ccgcaaagga cagetecagg gggttgattt tggaatggtg gaggcgctcg acaattgtca ggctttgaga cgagaactgc aatgcccagg aatgagaaag caaacacaa cgcagcttcg acgcggagtg ttggcccggg cgcaagaatg gagcctggtg tttgtgcttt agogtacctg gtggagctta aacaaaact cgagacette gatgttacat cttacagcct cctgctgctg gaagaattga 5 10 2

SEQ ID NO:50

atggetetet cactetteae tgttggacaa ttaattttet tattttggae acteagaate actgaagcc 25

SEQ ID NO:51

caaacgcctt atgaacaaa ctagcaattc tegetateat egetatggta etttteageg ggaagettet gecagag cagactccaa agcagattga gatcaaatat

SEQ ID NO:52

30

a tggtcagtgt gtgcaggctc ttgctggttg ctgccttgct gctgtgtttg caagcacagc gaatgcagct cttgcaagga gaaatcgact tgtctttctc tcagcactgg tctcatggct ggtaccctgg aggaaagaga tgaagcggga atctcaggcc actgaggacc aacatcctaa agagcatcct gattgadacc tctggggaga ttaaactgtg cctacagctc accagagata 35

aattccaaaa gaggaaatga

SEQ ID NO:53

121 ttgaaatgtg actttgctct aacatctccc atctctctgg atttcttttt gcctcattat gcgatgggga aaatcagcag caattattta agatctgcct ctgtgacttc ttgaagataa agatacacat gtcctcaaaa 1 tgtgttttgt agataaatgt gaggattttc tctaaatccc tcttctgctt gctaaatctc 61 actgtcgctg ctaaattcag agcagataga gcctgcgcaa tcgaaataaa caattcattt ccagactttg tacttcagaa tectgeecae tcttccaact catg 181 241 301 S

10

atgte aggeeegagg aegtgettet gtetaeegte ggetettgta etagtaetge tgagteteag cacttoggca ctaggg SEQ ID NO:54

15 SEQ ID NO:55

afsaggpgkg agefvhmarn epgrarnmts dietayvggs isavtrddlv yvdagpiyet sidrgvshfv vglladttlr rvtm grglgnsifc trapssvths tlqvlmgggg rleplelvnn svlhhdlprt sfnphlcnvl qpvgsagsfg snshfaavgv fthpdmfpvs igrqvlaqnq atresilygg lfgisascvp gemmhavafg lvskafvpss aafnasgeal afqhayssts tgvahaelvd psgirvatap hvqvafpvpp lesqvitved aglggnlfct dlhsrpdafi tkltdiketl hqlvipkppp thpnyeqtlt mspaaqlaka aarstcmtrl atesqvlktm nryrwmesca lsdeevarak nglkssllmn rratiafeae 1hpsrmvag rvaealvakp ptmvavgedl sslafksthg palteeeiae dtireyfaty mysrlytnvl

25 SEQ ID NO:56 (pTnMCS)

ccaegttege eggeateaga teggetateg gecattgeat aegttgtate catateataa atgcccagta tegetattae gattattgac tggagttccg ccccccatt atcatatgcc actcacgggg 61 ocgetacact tgccagegee ctagegeeeg etectteege tttetteeet teetteteg 1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga tgttgacatt agcccatata cccaacgacc gggactttcc cggttttggc agtacatcaa tgggcgtgga tagcggtttg catcaagtgt gtattagtca gcctggcatt tagtaatcaa ttacggggtc attagttcat atggcccgcc tggctgaccg catgtccaac attaccgcca ttcccatagt aacgccaata cttggcagta tcaatgacgg taaatggccc gtacatctac aaactgccca ctacttggca cttacggtaa tatattggct atgacgtatg tatttacggt tgggactttc cctattgacg taqttattaa cgttacataa gacgtcaata atgggtggag tatgtacatt aagtacgccc catgacctta catggtgatg 121 241 361 421 481 541 181 301 501 30 35

ctccaccca

atttccaagt

atacacccc gcttccttat gctataggtg atggtatagc ttagcctata ggtgtgggtt ttccattact aatccataac gccgcacaag cttcagagac gatctccacg cttgctccta tgttgtattc aattcacata agcttccaca cacggctgac cagtgtagtc ttttggcacc aaaatcaacg cctggagacg tccgcggccg gactaacaga caacagctta acttggccgt attgttaggt gcaaaacga ttatgagaaa ctttaaagtg aagaggaaaa acatgatatg gtaggcgtgt ccgcctatag actegatatt ctttatctgt agccgacctt gcatgctata ctgtttttgg cttggggcct tccaatctca ctcgacacgg ttgggacaag cttaggcatg gccatgggtt tatctattcq aatcgccctg cggactaggo ggctgcaacc aggggatcgc atactctgtc attatttaca catagcgtgg caaatgggcg tgccaagagt gacgtaagta gtagcggcgg teggeagete ncaccagtgt attgggctcg gcagctgagt cggtggaggg atagctgaca taaaacgact ctcttaccga gaatcgaccg ggcatgctag atattcgtga taagtogagt ccatgtgcga accaatttct gtgatgctgg tcagcaactt ctaaaagcaa cgatccagcc ctgttactct cgtcagatcg aaaatcagcg cggcaaagga acacccaaac aacttgttaa tcatgctgct gtcctgccta agaaacaagg ggaaattatg tgacgatact ctatatgcca ggtcccattt ttttattaaa ctcttctccg tcatggtcgc gagcgtggag caccgtcgga gaagatgcag gtgctgttaa accagacata tgggagtttg acaccgggac atgcccacca tttagtgaac gaattacact aacatcaaac gactggtctg agtegeacta caeggtegtt atgttcaaag aaagctcatg cccattgacg aaactctca gtataccgtt ctcattgtca tggtactggt tggaaaccta aagaggctga aaaggccgaa tactcagcgt cgttttgata acaagggaag gacttgaaaa gttcatgctc tacgetttgg aacqtactct ccctattgg tctctattgg cggacatggg ctccagcggc tacaggatgg tgcccgcagt gcacagcaca gtctgaaaat agcggcagaa tcccgttgcg იმიმინინი tttctgcagt gcagageteg attetgeece tccatagaag ggattccccg tggctcttat acttgactgt aacaaaacat gcgactcgct tgtacttgtt caccacaccg agcggaaaac tttaggctat ttgacgtcaa acaactccgc tgaaattcga gaagctgggt atctcgctct acctaaaatc aaccttccga cageteagag gcttgcgggc cagaaatcga ctacacaata cacacatggt gccacaacta attgaccact cagaggtaac totgtattt ccgtcccccg gtacgtgttc ggtcccatgc ccagacttag gacttaaggc tegttgetge attggaacgc tagggtatgt ccatgggtct ctctttacca ccaaagcgag tccacaaaga tttcagagca taccgagtaa agattgaaga aaatgtcgta tgttttgacc cacacccctt gccacgcatt tgcgagette gtctatataa cgatgcccat aatccgttga cagacctagg tattgtataa actcaaagac accacccgtc acttacctgt gccgaacgag taacatgttg ctaacacagt ggcattctgg aaaatttatt atggctcttt attgaccatt tgacacggac tacaacaacg acagtggagg ttacacgact acgttggctt cgaatctcgg tccgagccct gccgtggcgg gcagatggaa tgataagagt tgagcagtac ggactttcca acggtgggag ccatccacge ggaacggtgc actctatagg ctgttccttt aacctgccaa aatcgtcacc tegggcaata cttatggtat gegtteeege gcgagcattc ccatggtata gtacaatatg actcattgtc tgccaaattc ctagcaacta tcatctagtc aagcgaatgc atqcttcaac cacttccagg ctacgccata gaagttttgc ctactagete 1021 1081 1141 1201 1261 1381 1501 1321 1441 1561 1621 1681 .741 1801 1981 1921 781 841 901 961 1981 2041 2101 2281 2461 2521 2221 2341 2401 2161 2581 2641 2701 2761 2821 5 2 15 20 25 30 35

gccttgcagc gcccttccca gttaaaattc atgcatatcg aggccaattt attgctattt atatttcgcg ttcctgacgg ttagctcact tggaattgtg gcgcgcaatt tgacctcgag cactggccgt cggcaaaatc ttggaacaag ctatcagggc acattccqtc tttgaagcta teceetgage acgctaaacc cagcaaacag aagcgatccc taaacgctaa tttcgcccat gcaccacacc qttaaataaq acacaacata actcacatta gctgcattaa ttatgattga attactggta ggtaaaaatc ctctagaact ttaatgattt gtaaagcacc accaagaccc tccaagaagt tcctgtttac acgagaatta tegatacege ttaatattt aggccgaaat ccaccttaac atggctggtt ataggtttta gagggtcatt atacagtaaa ttataaataa agtgtaaaag gagtgagcta cagcaattaa tacgcgcgct aacttaatc cgcaccgatc ttgttccagt gaaaaaccgt aatgcaggta ccttcagata actgccgttt aaatggcatc tgtcgtgcca tctgccttag cttatcccta ccttqaatac acttaaagct taatgattt gtaatgtgag gtatgttgtg attacgccaa gtggcggccg aataatcct ttctgtaata tcacaattcc taattgcgcg agcacttgtc tgttatccgc ggtgcctaat atgaacttga aaaaaattag tgagtcgtat tggcgttacc gtcaaagggc agatgtgtat ctcagggtca ccgataaaaa aataaaatag gggttatcaa caggataata tagtgaatga aaggeteace aactaaccaa tggaatactg aataatcaac tagtgagggt tegggaaace gttggtgacc ctgttataag taataaaac aagagaacca atgaatccc cttccggctc tatgaccatg ctccaccgcg atcaagctta cgaagaggcc attgtaagcg tttaaccaat gggttgagtg aacttactca cactccctgt tccgagcagt tcatcacgat ctgatttaac aaatggttgg tatgatcccg ttgcattggt tgtgcttaaa aaacagggaa cageteattt gcgctgaaat ttgttccctt tgtgtgaaat cattgtcagc aaattgatcc ggaattcgat cgccctatag ggcgtaatag gcgaatggaa ggactccaac aaagagcttg gaaataacta ggatagacat atgctattca cgggaaagc ttattgataa ctgagcactg catcttgtca acactttatg aggaaacagc aaagctggag gggaaaaccc gaccgagata tcatatgaca ttcatcagtg ttgaaagata atogtaaaaa atgoootott tcatcgatag cccaaattgg ctgcaataaa cacaaaggct taaagcctgg cgctttccag tactcagtgc ttcttcctgc gccaaccatc tactccggga atgatgatga tttattgagc ggctatcaat cgatcagctt cttgattata attggttgga ggtggataca cccaggcttt aatttcacac ccgggctgca cattagggga gaaaaaccta atttggtgac aataaaatta tcatggtcat agctgtttcc gctcactgcc tccgggatct aaaaaacaat gtacccaatt cgtcgtgact ttcgccagct caaaagaata taaagaacgt ataacaccgg gcactactat gcataaagtg acaacaaaa ctcgttgacc aaagggaaca agcctgaatg tttgttaaat aacatgaagg atccagccat ctctttgttt tctagagcga tgcctacatc gatgcctatc aggtaaacgc tgatgtatac cgagccggaa tgaacattat aatgttaatt agcggataac agtggatccc gcgttaaatt agtccactat gatggcccac aggaaagctt aatcttcttt ttgaggggtt aaaggcaaat gtaatgaaaa gtgctggatt attgcgttgc ctttaaaaat cattaggcac aaccctcact 633366666 cgttttacaa acagttgcgc ccttataaat accgcggctt gaataacatc aataatccaa taatqqqcca accaccagcc attaagttaa acatccccct ttaccaaat caatacatgc ttgctgatga ttagtggcta 3421 3121 3181 3241 3301 3361 3481 3541 3601 3721 3901 5281 3661 3781 3841 4741 1801 4861 4981 5161 3961 4021 4081 4141 4201 4441 4501 1961 4681 4921 5041 5101 4261 4321 4381 4621 S 10 15 20 30 35 25

tcactcaaag ggcgctttct tcgtcttgag ctacggctac aaacccgaca tectgtteeg ttttgtttgc gctgggctgt caggattagc cggaaaaaga cttttctacg gagattatca aatctaaagt acctatctca gataactacg cccacgetca cagaagtggt tagagtaagt cgtggtgtca gcgagttaca cgttgtcaga ttctcttact taataccgcg gtcattctga acccaactga atttgaatgt aaggcaaaat cttcctttt ggcgctcttc cggtatcagc gaaagaacat tggcgtttt agaggtggcg tegtgegete cgggaagcgt ttcgctccaa ccactggtaa ggtggcctaa tcagtgaggc ccggtaacta cagttacctt geggtggttt recettegat ttttggtcat gttttaaatc ccgtcgtgta taccgcgaga gggccgagcg ctacaggcat aacgatcaag gccgggaagc cactgcataa gtcctccgat caatacggga actcaaccaa gttcttcggg ccactcgtgc caaaaacagg tactcatact geggataeat cccgaaagt ttgcgtattg ctgcggcgag gataacgcag cgctcaagtc ggaagctccc ctgctgaagc gccgcgttgc tttctccctt gtgtaggtcg tgcgccttat ttcttgaagt cgttaaggga ctggcagcag accgctggta tctcaagaag taaaaatgaa caatgettaa gcctgactcc gctgcaatga ccagccggaa attaattgtt gttgccattg tccggttccc ageteetteg gttatggcag actggtgagt tgccggcgt tegatetaac tgtctcatga attggaaaac tctgggtgag aaatgttgaa cgcacatttc ggtcgttcgg agaatcaggg ccgtaaaag caaaaatcga gtttcccct cetgteegee tctcagttcg gcccgaccgc tgctacagag tatctgcgct aaaaaagga cgaaaactca ccttttaaat cttatcgcca caaacaacc tgacagttac atccatagtt tggccccagt aataaaccag ttcattcagc atcactcatg catccagtct gcgcaacgtt ctttctgtg aaaagcggtt gagttgctct agtgctcatc gagatccagt caccagcgtt ggcgacacgg tcagggttat aggggttccg ggttatccac aggccaggaa acgagcatca ttaccggata aacgcgcggg ttacgcgcag tatttcgttc gctgtaggta taagacacga atgtaggcgg cagtatttgg ctcagtggaa gataccagge cccccgttca cttgatccgg tcacctagat aaacttggtc gcttaccatc atttatcagc tatccgcctc ttggtatggc ccgcagtgtt ccgtaagatg ttaatagttt tgttgtgcaa gaactttaaa cttttacttt tgcggcgacc taccgctgtt agggaataag ataaacaaat gaagcattta ctcactgact geggtaatac ggccagcaaa cgccccctg catageteae gtgcacgaac ggactataaa accetgeege tccaacccgg agagcgaggt gttggtagct actagaagga gcgatctgtc aagcagcaga gggtctgacg aaaaggatct atatatgagt atacgggagg ccggctccaq cctgcaactt agttcgccag agetegtegt gaatagtgta gccgcaaaaa tgatccccca agtaagttgg gtcatgccat ccacatagca tetteageat atttagaaaa tcaaggatct caatattatt 5401 5461 5581 5641 5761 5521 5701 5881 6001 1909 5821 5941 6121 6181 6241 6301 6361 6421 6481 6541 6601 6661 6721 6841 6901 7021 6781 6961 7081 7141 Ś 10 2 2 25 30

EQ ID NO:57 (prnMod)

35

CTGACGCGCC CTGTAGCGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG 50

1250 1350 1400 1450 1500 1100 1150 1200 1300 1550 1600 1650 1000 1050 1800 900 950 650 700 750 800 500 550 900 TCGTTGCTGC GGTCCCATGC ACAGTGGAGG GCCGCACAAG CIGITCCITI ACTIGATATI GCATGCTATA GCTATAGGTG ATTGACCACT CTTCAGAGAC ATTATTACA TTTTATAAA CGGACATGGG GAAGATGCAG TCCCGTTGCG AAACTGCCCA TCGCTATTAC ATTGGGCTCG CICCITICGC TATATTGGCT TAGTTATTAA TGGAGTTCCG CCCAACGACC AACGCCAATA CCTATTGACG CATGACCTTA TAGCGGTTTG **JGGGAGTTTG** ACAACTCCGC GTCTATATAA CCATCCACGC TCCGCGGCCG GACGTAAGTA ATGGCTCTTT CCATGTGTGA AATCCATAAC CCACCAGTGT TGAGCAGTAC GACTAACAGA CIAGCGCCCG AAATGTCGTA TGGCTCTTAT GCTTCCTTAT ATTGACCATT ATACTCTGTC TGCCCGCAGT TCCGAGCCCT CITGCICCIA GAGCGTGGAG AGCGGCAGAA CAGAGGTAAC TATGTACATT GATTATTGAC TGGCTGACCG TTCCCATAGT TATTTACGGI AAGTACGCCC ATGCCCAGTA GTATTAGTCA TIGACGICAA ACGGTGGGAG CCTGGAGACG CGATCCAGCC TGCCAAGAGT GGTCCCALTT GIACGIGITC AGCCCATATA TGGGCGTGGA CGGCATCAGA GTAGGCGTGT TTCCATTACT GTCTGAAAAT GACTTAAGGC TGATAAGAGT CAGTGTAGTC ATAGCTGACA CACCGTCGGA TGCCAGCGCC CCACGTTCGC CATATCATAA ATGACGTATG ATGGGTGGAG ATCATATGCC CTCCACCCCA GGACTTTCCA CGTCAGATCG ACACCGGGAC GGATTCCCCG GGTGTGGGTT TACAGGATGG CCGTCCCCCG AGCTTCCACA ATGCCCACCA TGTTGACATT ATTAGTTCAT ATGGCCCGCC GCCTGGCATT GTACATCTAC AGTACATCAA CACACCCCTT ATACACCCC CTATATGCCA CGAATCTCGG TCGGCAGCTC ACCAGACATA ATTACCGCCA TTACGGGGTC ATTGACGTCA AAAATCAACG CAAATGGGCG TTTAGTGAAC TCCATAGAAG ATTGGAACGC TACAACAACG GATCTCCACG CGGTGGAGGG CTTACGGTAA CATCAAGTGT CTACTTGGCA CGGTTTTGGC ACTCTATAGG CTTGGGGCCT TGACGATACT TCATGGTCGC GCACAGCACA TAGGGTATGT GCAGATGGAA TGTTGTATTC TTTCTGCAGT CCGCTACACT TCCTTTCTCG ACGITGIAIC GACGTCAATA TAAATGGCCC ATTTCCAAGT TTAGCCTATA TCTCTATTGG TCTGTATTTT GTAGCGGCGG GGAACGGTGC GCCATTGCAT TGITTIGACC CCCCCTATAG CATGICCAAC TAGTAATCAA ACTCACGGGG TTTTGGCACC CCCATTGACG GCAGAGCTCG CTGTTTTGG ATGGTATAGC CCCCTATTGG TGACACGGAC AATTCACATA CATAGCGTGG CTCTTCTCCG CTCCAGCGGC CCAGACTTAG GCCGTGGCGG CACGGCTGAC GCAGCTGAGT GTGCTGTTAA 20000000000 CCATGGGTCT TTTCTTCCCT CCCGCCCALT TGGGACTTTC GCCACAACTA CGCAGCGTGA CGTTACATAA GGGACTTTCC TCAATGACGG CATGGTGATG CTTGGCAGTA 30 35 S 2 25 10 15

2100 2150 2200 2250 2350 2500 2300 2400 2450 2550 2600 2650 2700 2750 2800 2850 2900 2950 3050 3100 3000 3150 3200 3250 3300 3350 3400 AAAACTCTCA TAAAACGACT TCCACAAAGA AACAAAACAT TCGGGAATAC CAAAAACGAC TGTTCAAAGA AAGCTCATGA ACCACACCGC TGTTACTCTT ATCCGTTGAG TACAATATGC CATGATATGT TAAAAGCAAT CCTAAAATCT AGCGAATGCA AAGGCCGAAA CTTACCTGTT GGACTAGGCC CATGCTGCTA TTCATGCTCA AGAAATCGAA GCATTCTGGC TACTAGCTCA TAATGATCCA TGTGTGTTGG TGTTGTTTGC CCACTGTCCT GGATTGGGAA GTACCTCTCT AGGTGTCATT CICICICICI TGCTGAAGAA TTAGAAAGTA TIGAIGCCIA GAATTACACT ACTTGACTGT CCAAAGCGAG AATCGTCACC CTTTATCTGT TATTCGTGAG ACGGTCGTTC ACCGAGTAAC CATGGTATAA AGAGGAAAAG AGAGGCTGAC ATCTATTCGA CAGCAACTTA TCTCGCTCTA CCACCCGTCA TAGCAACTAA TCCTGCCTAC CTTGCGGGCG TAACACAGIC GAAATTATGA GTTTTGATAT AAGTTTTGCG GCTGCAACCC TCTAGAGATC TTGTCTGAGT GCAAGGGGGA GGCTCTATGG GCCAGCCATC GGTGCCACTC CTCTCGGTAC CGGTACCAGG ATCACTTTAA TIGGICIGCC ATTCTGCCC GTCGCACTAC GCCACGCATT ATTGTTAGGT GGCATGCTAG ACTGGTCTGA TTCAGAGCAA AACCTGCCAA CGAGCATTCT GGAAACCTAT TTTAAAGTGC AAGTCGAGTA CTCATTGTCA CCATGGGTTC TTAGGCATGG TTAGGCTATA ATTGTATAAA ACTIGITAAT ACTTGAAAAG AGCTCAGAGC AACATGTTGG ACTICCAGGC CTTACTCGTG ACGCTTTGGG AGATCAGAGC CCTTCTAGTT GACCCTGGAA TTGCATCGCA GGGCAGCACA GGATGCGGTG CTCTCTCTCT GCCTTTTATC TAACAAATGG CTCTTTACCA GAATCGACCG ACGITGGCTT GTACTTGTTG ACTTGGCCGT GTATACCGIT GCGAGCTTCA CGTTCCCGCT GCCGACCTTG TGATGCTGGC GGTACTGGTT GCGGAAAACT CTCAAAGACT GCCAAATTCT TCGACACGGA CACCCAAACA ACCTTCCGAG CCGAACGAGC GGCAAAGGAG TGCTTCAACT TGGGACAAGC AACAGTTCGC CAAGGGAAGA GGCTAATAAA ACACATGGTT ATCTGCTGTG TGCCTTCCTT AATGAGGAAA GGGTGGGGTG GGCATGCTGG CTCTCTCTCT AAAACTGATT GACCAAAGGT TTACATGATT CAACAGCTTA CTCTTACCGA AACATCAAAC GCGACTCGCT TCATTGTCAG GATGCCCATT TTAIGGIAIL TATGAGAAAG CCAATTTCTA AAGCTGGGTT AGACCTAGGA CATCTAGTCA CCAATCTCAT AAATCAGCGC ACTCAGCGTC GATTGAAGAA GAAATTCGAA TACGCCATAG ATCGCCCTGA GAAACAAGGT ACGTACTCTC AAATTTATTC GATCACTTCT TTTTTG TACACAATAA CCTCCCCCG TTCCTAATAA CTATTCTGGG SACAATAGCA TUTCICICT CTCTCTCTCT CATCACAACA FTGACCCGGT CAATTACTCA 5 10 15 20 25 30 35

4450 4500 4550 4700 4850 5150 4050 4100 4150 4200 4250 4300 4350 4400 4600 4650 4750 4800 4900 4950 5050 5100 5200 3750 3800 3850 3900 3950 4000 5000 ATCGCAATAC ATAGATAGGT GGTTAACATG ATTAAAACAG CGAGGGGGG CGCTCACTGG GACAAGATGT GGGATTCATC GCTTATGATG ATTTATTGCT TGACGAAATA AACCAATAAT ACATCACTCC AAATGCACTA CCATTTAGTG AAACCTTATC TTGAAAAAA CAGCAAATTG AATTCTCGTT AATCATTAAG TGAGTTAGCT CAGCTATGAC TCCCCCGGGC **LACCCAACTT** ATAGCGAAGA AATGGCGAAT AATTTTGTT AATCCCTTAT CAGTTTGGAA GGGCGAAAAA TCTTGGGTTA ATAACAGCAA CACCAATAAT AACAAAAGCT GCTCGTATGT GGTTATGCAT AAAAAGGCCA TAAAACGCTA GTTTGGATAG CTAAAAAGGC GTTTTTCGC TATATTE ATAATAA AGCTGGCGTA GCGCAGCCTG ATTCGCGTTA AAATCGGCAA AAATCATTAG GATAAATAAA CATCATTTGG GAGCTTGAGG ATGATTATAA TGGTAAGGCT AGCCAATAAA TGGAATGAAC ACGCCATTGT ACGGAATGTT TTTTGGTAAA CCCGGTAATG TAIGCTICCG ACACAGGAAA CACTAAAGGG AACTAGTGGA CCGCTGACCT GIATTACGCG ACCCTGGCGT AGTGTTGTTC GGGATCATAT CGTCAGGAAA AAAAATGCCC CAACGICAAA CTTTATCGTA ATTAACATTC CITIGCCGATA AATAATACAG CCGGTTGCAT TCCCACCACC TATCATTGGT AGCITICCIG CCCCTAATGA GTCATATGAT CTTTACACTT TAACAATTTC AATTAACCCT GCCGCTCTAG CTTATCGATA GACTGGGAAA CCCTTTCGCC CCCAACAGTT TTTTGTTAAA CAATAGGCCG GATAGGGTTG CCACTACTCC CTCAATGGCT GAGCTTGAAA ATGACTCTTT GATATAAACG CAGTACTGCC GGTAAGGTAA ATAGTGAGTC ACGTGGACTC ATTTTACCA CGCGGAATAA TTACTCCCT TTGGTAGTGA ACTGATGAAT CCAAGCGCGC AATTCGCCCT CTATTAAAGA TAAATCCGAG TTATCTTGAT ACCAACTTAA TGTGAGCGGA CGCGGTGGCG GATCGCCCTT AGCGTTAATA ATTTTTAAC AATAGACCGA GTCAACGAGA GCCAATAACA CACCTTGCTG GGTAAAGCGA CCAACCTTCA AAGTGATGCC ATACATTACT TACACATCTT GCACCCCAGG CGATATCAAG ACAACGICGI CAGCACATCC GGGCGATGGC TAACTTAATG TAAAAACTTA CCTAAAAGAG GCTTTTTATT GCTAAATCTT CATTATATT TGTCTCCTGT ATAGCAGGAT CCATCCCAAA AAGGTCATCG TATTTGAACA CCTATCCAAG TTAGCCTTGA ATCCAAGAGA GACCCTGAGC TTAAGGTGGA TGTGTGGAAT GGAGCTCCAC TGCAGGAATT CCCGGTACCC AAATCAGCTC AAATCAAAAG ATTTACCGCG TTTATTGAA ACTAAGCACT CCAAATCCAG ACAGTAATGG CCCTGTAAAG CTGTAATGCA GGAAAACTAA CTATCTGCAA CACTCATTAG CATGATTACG AATCGCCTTG GGCCCGCACC GGAAATTGTA CAAGAGTCCA CCGTCTATCA AGTGCTCAGG ATGCGAAAA TCAAGAGGGT CCGTCGTTTT GTATCCACCT ATGATGTGCT 2 S 30 35 15 20 25

5500 5600 5700 5550 5650 5750 5800 5850 5900 5950 6000 6050 6100 6150 6200 6250 6300 6400 6350 6450 6500 6550 0099 6650 6700 6750 6800 6850 6900 6950 AAAGACCAAG CGATTTCTGT AAATAATAAT GCTTTTGTTC TCATAGCTGT CATACGAGCC GCTAACTCAC AACCTGTCGT CGGTTTGCGT ATACGGTTAT AAAAGGCCAG GCICGGICGI TTTCCATAG AGTCAGAGGT TCACGCTGTA CTGTGTGCAC CCCTGGAAGC GATACCTGTC ACTATCGTCT GCAGCCACTG AGAGTTCTTG TIGGIAICIG AGCTCTTGAT TTGCAAGCAG TGATCTTTTC GGGATTTTGG AAATTAAAA GGTCTGACAG TGTCTATTTC TACGATACGG GAGACCCACG GGAAGGGCCG GTCTATTAAT TCGTTTGGTA ACTGAGTGTA TTCATCATCA CAATGCGCTG ATACCGATCA GTAATCATGG TTCCACACAA TAATGAGTGA CCAGTCGGGA CGGGGAGAGG GACTCGCTGC AAAGGCGGTA ACATGTGAGC TTGCTGGCGT TCGACGCTCA AGGCGTTTCC CCGCTTACCG TTCTCATAGC CCAAGCTGGG TTATCCGGTA GCCACTGGCA GCGGTGCTAC AGGACAGTAT AAGAGTTGGT GTTTTTTGT GAAGATCCTT CTCACGTTAA AGAICCITIT CTCAGCGATC TGTAGATAAC GICACGCICG ATGATACCGC CCTCCATCCA CCAGITAAIA GAGTAAACTT CCAGCCAGCC GGCTTGGAAT TAAGTGATGT CATCATGCTA GATTGGCTAT CCGCTCACAA CGCGCTTGGC CIGGGGTGCC GGCCAACGCG TGCCCGCTTT CTCGCTCACT CAGCTCACTC GCAGGAAAGA AAAGGCCGCG ATCACAAAAA TAAAGATACC TCCGACCCTG GCGTGGCGCT ACGACTTATC AGGTATGTAG CTACACTAGA GGTAGCGGTG GICGLICGCI CCGCTGCGCC CCTTCGGAAA AGGATCTCAA GGAACGAAAA ATCTTCACCT AAGTATATAT CICCCGICG CAGTGCTGCA AGGCACCTAT CAGCAATAAA ACTITATCCG AAGTAGTTCG GCATCGTGGT CTGCCACAAA AAAAGCCAAC CACCGTGCTG CATCGTTAAA GGGTTAATTG TIGCGCICAC AAATTGTTAT AGTGTAAAGC TTAATGAATC AGGGGATAAC GGAACCGTAA CTTCCGCTTC CGAGCGGTAT CCTGACGAGC GACAGGACTA GCICICCIGI CCTTCGGGAA TTCGGTGTAG CCGGTAAGAC TTCAGCCCGA TAGCAGAGCG CTAACTACGG AAGCCAGTIA GCAGAAAAA TTAATCAGTG AACCACCGCT GACGCTCAGT ATCAAAAAGG AATCAATCTA AGTIGCCIGA CATCTGGCCC CCAGATTTAT TGGTCCTGCA AAGCTAGAGT ATTGCTACAG ACCCGTAATG SCIAITCIIC ATTAATTGCG CAACAAATGG CCTTTAGTGA GGAAGCATAA GCCAGCTGCA CCACAGAATC GCTCCGCCCC AATAGCACCA TTCCTGTGTG ATTGGGCGCT CAAAAGGCCA GGCGAAACCC TCCCTCGTGC TCGGCTGCGG GGTATCTCAG GAACCCCCCG CGCCTTTCTC TGAGTCCAAC GTAACAGGAT AAGTGGTGGC CGCTCTGCTG CCGGCAAACA CAGATTACGC TACGGGGTCT **ICATGAGATT** TGAAGTTTTA TTACCAATGC GITCAICCAT SAGGGCTTAC AGCGCAGAAG retteccese CTCACCGGCT CGTTGTTGCC 5 10 15 2 25 30 35

	TGGCTTCATT	CAGCTCCGGT	TCCCAACGAT	TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGCGAGT TACAIGAICC / 100	TACATGATCC	007/
	CCCATGTTGT	GCAAAAAAGC	GGTTAGCTCC	CCCATGITGI GCAAAAAGC GGITAGCICC ITCGGICCIC CGAICGIIGI 7150	CGATCGTTGT	7150
	CAGAAGTAAG	TTGGCCGCAG	TGTTATCACT	CAGAAGIAAG TIGGCCGCAG IGITAICACI CAIGGIIAIG GCAGCACIGC 7200	GCAGCACTGC	7200
	ATAATTCTCT	TACTGTCATG	CCATCCGTAA	ATAATICICI TACTGICATG CCATCCGIAA GAIGCITITC IGIGACIGGI 7250	TGTGACTGGT	7250
ς.	GAGTACTCAA	CCAAGICATI	CTGAGAATAG	GAGTACTCAA CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTTG 7300	GACCGAGTTG	7300
,	CTCTTGCCCG	GCGTCAATAC	GGGATAATAC	CICITGCCCG GCGICAAIAC GGGAIAAIAC CGCGCCACAI AGCAGAACII 7350	AGCAGAACTT	7350
	TAAAAGTGCT	CATCATTGGA	AAACGIICIT	TAAAAGTGCT CATCATTGGA AAACGTTCTT CGGGGCGAAA ACTCTCAAGG 7400	ACTCTCAAGG	7400
	ATCTTACCGC	TGTTGAGATC	CAGTTCGATG	ATCTTACCGC IGINGAGAIC CAGITCGAIG TAACCCACIC GIGCACCCAA 7450	GTGCACCCAA	7450
	CTGATCTTCA	GCATCTTTTA	CTTTCACCAG	CIGATOTICA GCATOTITIA CITICACCAG CGITICIGGG IGAGCAAAA 7500	TGAGCAAAAA	7500
10	CAGGAAGGCA	AAATGCCGCA	AAAAAGGGAA	CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAAATGT 7550	ACGGAAATGT	7550
) 	TGAATACTCA	TACTCTTCCT	TTTTCAATAT	TGAATACTCA TACTCTTCCT TITTCAATAT TATTGAAGCA TITATCAGGG 7600	TTTATCAGGG	7600
	TTATTGTCTC	ATGAGCGGAT	ACATATTIGA	TIAITIGICIC AIGAGCGGAI ACATATITGA AIGIAITITAG AAAAAIAAAC 7650	AAAAATAAAC	7650
	AAATAGGGGT	AAATAGGGGT TCCGCGCACA TTTCCCCGAA AAGTGCCAC	TTTCCCCGAA	AAGTGCCAC		7689

SEQ ID NO:58 (a Kozak sequence) ACCATGG 15

SEQ ID NO:59 (a Kozak sequence)

ACCATGT

20

SEQ ID NO:60 (a Kozak sequence) AAGATGT

SEQ ID NO:61 (a Kozak sequence)

ACGATGA

25

SEQ ID NO:62 (a Kozak sequence) AAGATGG

SEQ ID NO:63 (a Kozak sequence) 30

GACATGA

(a Kozak sequence) SEQ ID NO:64 ACCATGA

(a Kozak sequence) SEQ ID NO:65 5

ACCATGT

SEQ ID NO:66 (conalbumin polyA)

accaacttgg cttcgctact ttcaagagtc aatgtggctt cctgaggttt agtgccaggc aaagcaaaat gctttcgctt cgtcactctg gctgcaggcc aatatttcc aaatgggttt attatctagt etgetteete tgeeetteet aatccagttc ctcaacatct agcaccctt cctgattaac aggcgtttgt gaaataaaat ttccttcccc ccatgagaga atcccattgc ctccc tctgccattg gttaagcaca ctactgctgc tgccctagat gccacagcaa 10

(synthetic polyA) SEQ ID NO:67 15

**3GCGCCTGGATCCAGATCACTTCTGGCTAATAAAAGATCAGAGCTCTAGAGATCTGTGTGTTTGGTTTTTTT** IGGAICIGCIGIGCCTICIAGIIGCCAGCCAICIGTIGIIIIGCCCCCICCCCGIGCCIICCTIGACCCIGG **AAGGTGCCACTCCCACTGTCCTTTCCTAATAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAT** ICTATTCTGGGGGGGGGGGGGGGGGGGAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGG 20

(avian optimized polyA) SEQ ID NO:68

tttatcatca ctttaaaaat aaaaaacaat tactcagtgc ctgttataag cagcaattaa tctqccttaq cttatcccta ccttgaatac tccgggatct cgggaaagc gttggtgacc aaaggtgcct aaatggttgg taataaaac aaaaaattag aggtaaacgo cattgtcago aaattgatco aagagaacca ctgatttaac ttattgataa gatgcctatc attggttgga atgaacttga acaacaaaa cttgattata tgcctacatc tgaacattat ggggatcgc tctagagcga ttatgattga aaagtatatt attactggta tccaagaagt 25

30

ID NO:69 (vitellogenin promoter)

TGCAGGTTTT GATGAAGTTG CTGTAGGAAA TGAGAATCTC ATATTTACAG TGAATGTGTT CTTGTGTTAT CAATAAAT CACAGTTAGT GTATITGGII CCIGCACTIC TAICCCICIT GCCITACIGC AATIGITCAC ATTITGCTCC CATITACITI GGAAGATAAA GCTGCATTTT CAGCTACTTG TGCATCAGIT TCTAGGCTGA 35

GATTACGCTC TCAGGTGCCA GCCTCTTAAA AGTCAATGTA CCAGCAGCCA GTAAATGCCG ATTTATGATT AGAACACATT AAGGTTTTTT CGTGAAGAGA CGGAGCTGAA GGTTTATCTG GGGAGAGTAT ACAACTTGGG GGACCATGTA TAATTTATT CAATCAGAAA AGGAATTCAG GCCTGGCAGA TTTCTCAGTA AAAAAACTTG CTGAACATIC TICCATAAAA GICICACCAI GAATAGCAAA CAGCTAGATA ATCTGTGGCA ACATCAATTT AACTICIGGT CTTTATTCGC TCAGCGTGAC TGAGGGAAGA CCATATATTG GAGTGTTTAC ATAAACTGAT AGCAAGTAGC GATACATTGA ATTIGATITI TCATTTAAAA ATATTCCTGG ATGAATACAT GAATGGCCAT AAACCAAAGC GGGTTATTAT AACCAGTCTC TACATATGTT TGGTGCACAT TTTCAATAAA ATTGCTGATT ACCTGGGCAA AATCTAGGAA CCAAGGTATT GTTTTACTGT GATTATACTG GGTCAACATA GCCGTGACCC GATCCCGTGA TCATGTGCGT TGGCCTGCAG ATCAGAGATG GCAAAAAGAG CCACGIGITC AAACCTTTGT TAGTAGAAGT CCTTCGCT S 9

GAGGICAGAAT GGITICITIA CIGITIGICA AIICIAITAI IICAAIACAG ID NO:70 (fragment of ovalbumin promoter - chicken) GCATTGTATG AACAATAGCT TCTATAACTG AAATATATTT GCTATTGTAT ATTATGATTG GAACACTGCA TAAAGCATTC AAAGAAAGTG ACAAGAAGGA TAATAAGCAC AAGCTCAATG AGTCCTGATG TICCICIGIC GCATGCTTCT TTGTTCTCAC TTCCCACATT TAAAAACTAA TAAAATCTAA CCTCCAGCTG AATTTCACAA CAGCTATAGG GTTGTAACTG CAAAGGTCAA CCATTAAGIT ATTCATGGAA GATCTITGAG AAGTTTGTTA TCCTTCACAT AGGTCACGTC CGTTGTAGAT CCCATCCAAC TACCCAGAAT CTATTGAAAC ATATATTCCC GTATTGTTTT ACAACCATAG ATCTCTAAGG TTGAAATTGA AGAAAATAAT GGTGGGTCAC AATTCAGGCT TTAAATATTC CTCAGTTGGC CAAAATGGAC TCTATGGCGT CTCAGAAAAA CTCAGATGCA TATGTTTCC CCATTATCTG AGTCTTCTCT AACACATTGT GTCTGTTACT CATGAACACT ATAAACACAT AAAGATAGAT TATTTTGTCA CCTAGCATGG AATTGCTAAT ACAGGCAGAA CCATTCAATC TTAAATGATT TTGCTGTATC CAGACTTCTG AATATTTCCC TCACTCTAGT CTTAAAGATC TCCCTCGAAC AGTTCATATC ATCTGCCAGG ACACCCATAA TTATGICCIG ACTAACTAAT GATTAGCAGA TATTIGCICT CCCAATCCCA GGAACCTGTG GAGCTATGTT CGTCTGCTCT TTATTTCTCC TCAAATGAAA GAACATGAGC TTTCTGTTTT 15 2 25 30

SEQ ID NO:71 (chicken ovalbumin ehancer)

caaaacaqct tettteecte acaattaatg caagttctga cttcatatcc atgagactgt agagccttag ttgtactgct cacaatccca gcaatggcaa aatggttgct caaacagaga ctcacctgga gtctttattc aaagggcagc tcccaggtaa aacaaaatct acaaaactat ggggaaacaa accaaggtat tctctccatg caaagaggag gacctgtgag aggtctccga ctggcatagg acaatgtgtg ttgccaccta gcaacaaga tggctacagc aaatgattta gcattgtcaa tcttcaaact attgtagtgg aagaggcttt gggcaatcca tataactttc aatgaaatgc ctcacaatat tctctctaga tgttgcttac gacctgagca taatgtcaaa tacctgattt ccatcactga aatttccaca gtgattggat aaagtgcttt acagacccac cttgggacaa gtgctgggca cttgacctga cagagagaa attcatctgt atgaaaaggc tgctccttcc gtaggtttta ttttggataa tggtttaggg ttgcacagct gcaagaagat stgacettt S 10

GTGGATCAACATACAGCTAGAAAGCTGTATTGCCTTTAGCACTCAAGGCTCAAAAGACAACTCAGAGTTCAC untranslated region) (2, ID NO:72 SEQ ט

SEQ ID NO:73 (putative cap site)
ACATACAGCTAG AAAGCTGTAT TGCCTTTAGC ACTCAAGCTC AAAAGACAAC TCAGAGTTCA

GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG TATTCAAGGA GCTCAAAGTC TCTAGCCATG TCGCTTTGAT AAACGTTCAC TICGLICAGO CTTGCAGTGT TGCAGATCAA CAGAAATGTC GACAATGAGC TTTCAGAGTG TAGAGTGGCA GATCAAAGTG CALTGICTIC CACCAIGCCA AIGAGAACAI CIICIACIGC CCCAIIGCCA ICAIGICAGC GCACATCTGT ATGATGTTTA TGCCAGAATA TTCAAACAGC ATGGAATTAT AAGCAATGCC ATAAGGTTGT TGGTTAATGC TIGGITIALL TTGCCAGTGG AAGAGAGGAA CATCTGTCTT (Chicken Ovalbumin Signal Sequence) ACACAGATAA GCTCAGTGTG ACCAAACCAA TACCCAATCC CCTATCAACT AGTCAGACAA GCAATGGTTC GAAGACACAC ATGTACCAGA GGCCTTGAGC AATGTTATGG GAGCTTCCAT TACAACCICA CAGCACCAGG CAGTATTGAA CAACCAAATC TGAAGAGAGA AGGCTTGGAA CTGGGTAGAA TTCTCAAACT TGTGCAGATG GGAGGAAAAA ATTTAAGGAT GAAGATCCTG GACCAGITCI TGAAGTCTCA GTGCAAAAGA GATTCGGAGA GAGACATCCT GACTTTATGC TGTATAGAGG TCATCAATTC GCTCCGTGGA GGGAGAAAAC AAAGCAAACC CTGAGAAAT TGTTGCCTGA TGACTGAATG GCATGAAGAT ID NO:74 GTATACCTGG AAACTTCCAG TCTTCACTTA CTTGCCAGIA GTGAAGGAAC GCCAGAGAGC ACTGAGCAAG CTTCAGCCAA AAAGGACTGT TCAATGGCTT ATGITGGIGC TACTTACCTC 2 25 30 35

AGGCAGAGAG AGCAGAGAGC TCAATGAAGC GCATCTCCTC TCTCTGAAGA TCCTCTTCTG TATCAAGCAC ATCGCAACCA ACGCCGTTCT CATGCAGAAA TGGAGTGGAT GCTGCAAGCG CICITICAGCC AATCIGICIG CCATGCAGCA CAGCAGAGGC GGCATTACTG ACGTGTTTAG CTCAAGCTGT GTGGTAGGGT GACCATCCAT CTGAAGATAT

AGATGTGTTT CCCCT

S

ID NO:75 (Chicken Ovalbumin Signal Sequence-shortened approx. 50bp) SEQ ID NO:75 (Chicken Ovalbumin Signal Sequence-shortened ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG TATTCAAG

GCTCAAAGTC approx. Sequence-shortened TATTCAAGGA GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA JC Signal (Chicken Ovalbumin NO:76 日 100bp) 10

15 SEQ ID NO:77 (vitellogenin targeting sequence)
ATGAGGGGGATCATACTGCATTAGTGCTCACCCTTGTAGGCAGCCAGAAGTTTGACATTGGT

SEQ ID NO:78 (p146 protein)

KYKKALKKLAKLL

20 ATAMAKAMAKAMA

SEQ ID NO:79 (p146 coding sequence)
AAATACAAAAAGCACTGAAAAAACTGGCAAAACTGCTG

SEQ ID NO:80 (pro-insulin sequence)

TTIGTGAACCAACACCTGTGCGGCTCACACCTGGTGGAAGCTCTTACCTAGTGTGCGGGGAACGAGGCTT CTTCTACACACCCAAGACCCGGGGAGGCAGAGGACCTGCAGGTGGGGCAGGTGGAGCTGGGCGGGGGGCC CTGGTGCAGGCAGCCTGCAGCCCTTGGCCCTGGAGGGGTCCCTGCAGAAGCGTGGCATTGTGGAACAATGC TGTACCAGCATCTGCTCCTCTACCAGCTGGAGAACTCTGCAACTAG 25

30 SEQ ID NO:81 (spacer)

(GPGG) x

SEQ ID NO:82 (spacer)

GPGGGPGGGPGG

SEQ ID NO:83 (spacer)

SEQ ID NO:84 (spacer)

5

SEQ ID NO:85 (repeat domain in TAG spacer sequence) Pro Ala Asp Asp Ala

Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala SEQ ID NO:86 (TAG spacer sequence) Pro Ala Asp Asp Ala Pro Ala Asp Asp 10

SEQ ID NO:87 (9P41 epitope)

15 Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu

SEQ ID NO:88 (polynucleotide sequence encoding gp41 epitope)

Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu Asp Asp Asp Lys

SEQ ID NO:89 (enterokinase cleavage site)

DDDDK

20

SEQ ID NO:90 (TAG sequence)

Pro Ala Asp Asp Ala 25

Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys lle Leu Lys Gly Ser Cys Gly Trp

Ile Gly Leu Leu Asp Asp Asp Asp Lys

SEQ ID NO:91 (altered transposase Hef forward primer)
30 ATCTCGAGACCATGTGAACTTGATATTTTACATGATTCTCTTTACC

	SEQ ID NO:92 (altered transposase Her reverse primer) GATTGATCATTATCATAATTTCCCCAAAGCGTAACC
2	SEQ ID NO:93 (Xho I restriction site) CTCGAG
5	SEQ ID NO:94 (Bcl I restriction site) TGATCA
2	SEQ ID NO:95 (CMVf-NgoM IV primer) TIGCCGGCAICAGAITGGCIAT
15	SEQ ID NO:96 (Syn-polyAr-BstE II primer) AGAGGTCACCGGGTCAATTCTTCAGCACCTGGTA
	<pre>SEQ ID NO:97 (pTnMOD (CMV-prepro-HCPro-Lys-CPA) 1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga 61 ccgctacact tgccagcgcc ctagcgcccg ctcctttcgc tttcttccct tccttctcg</pre>
70	oggcatcaga tatattggct taqtaatcaa
25	ogttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt aagtacgcc cctattgacg tcaatgacgg taaatggcc gcctggcatt
30	541 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 601 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 661 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 781 acggtggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
35	tgttttgacc attggaacgc cacacccctt gcttccttat attgaccact

aattcacata gatctccacg cttgctccta gccgcacaag cacggctgac tgttgtattc cagtgtagtc gactaacaga acttgatatt caacagetta attgttaggt ctttatctgt qcaaaacga ttatgagaaa agccgacctt ctttaaagtg aagaggaaaa tccaatctca gccatgggtt acatgatatg ctcgacacgg tatctaftcg cggactaggc aatcgccctg ttgggacaag cttaggcatg acttggccgt ataatgatcc gttttttgtg ggtacctctc gtgaattaat ggctgcaacc attgcatcgc agcaaggggg tototototo tgaccaaagg ataaqcaqca attatttaca ccaccagtgt catagcgtgg toggcagete attgggctcg gcagctgagt cggtggaggg atagetgaca ccatgtgtga taaaacgact ctcttaccga gaatcgaccg ggcatgctag atattcgtga ctgttactct gtgatgctgg aaaatcagcg taagtcgagt tcagcaactt ctaaaagcaa aacttgttaa accaatttct cggcaaagga gtcctgccta tcatgctgct agaaacaagg caacagttcg gggctctatg acttactcgt ggaaattatg ctgtgtgttg ccctcccc aaatgaggaa ggggcagcac cctctctctc agtgcctgtt ttaacaaatg attgacccgg ggtcccattt ttttattaaa ctcttctccg tcatggtcgc atgcccacca gagcgtggag gaagatgcag accagacata caccgtcgga gtgctgttaa gtatacagtt gaattacact aaaactctca aacatcaaac gactggtctg cacggtcgtt aaagctcatg ctcattgtca tggtactggt tggaaaccta aagaggctga aaaggccgaa acacccaaac cgttttgata gttcatgctc gacttgaaaa tactcagcgt aacgtactct acaagggaag ctctagagat ctgttgtttg tctctcggta aaaaactgat tacgetttgg tttcctaata 9999t9999t gggatgcggt gtgctgaaga acaattactc tgcccgcagt ctccagcggc gcacagcaca gtctgaaaat tacaggatgg tcccgttgcg tttctgcagt attctgcccc aacaaaacat tgtacttgtt cggacatggg agcggcagaa cacacacacc acttgactgt gegacteget agtogoacta atgttcaaag caccacaccg gaagctgggt agcggaaac aaccttccga tttaggctat acctaaaatc tgaaattcga cageteagag atctcgctct gcttgcgggc cagaaatcga ctacacaata cacacatggt aagatcagag tqccaqccat ccactgtcc acatcacaac tctattctgg aggcatgctg tototototo teggtaceag aaaataaaa ccgtcccccg ccagacttag tagggtatgt gacttaaggc tegttgetge totgtatttt gtacgtgttc ggtcccatgc cagaggtaac ccatgggtct ctctttacca gccacgcatt ccaaagcgag cgatgcccat tgcgagcttc tttcagagca taccgagtaa agattgaaga tccacaaga aatccgttga tattgtataa accaccogto acttacctgt cagacctagg actcaaagac gccgaacgag taacatgttg ggcattctgg ctaacacagt aaaatttatt tggctaataa gccttctagt aggtgccact taggtgtcat agacaatagc tetetetete tetetetete catcacttta attgatgcct tgacacggac acagtggagg tgagcagtac tacaacaacg gcagatggaa ctattccttt aatcgtcacc tegggcaata cqaatctcgg tecgageeet gccgtggcgg tgataagagt ttacatgatt acgttggctt aacctgccaa cttatggtat gegtteeege gcgagcattc ccatggtata gtacaatatg tgccaaattc actcattgtc tcatctagtc ctagcaacta aagcgaatgc atgetteaac ctacgccata gaagttttgc ctactagctc cacttccagg agatcacttc gatctgctgt tgaccctgga attgtctgag aggattggga tctctctctc tetetetete tgccttttat attaattatg 1381 1441 1501 1621 1561 1891 1741 1861 1801 1921 1981 2101 2161 2041 2221 2281 2341 2401 2641 2701 2821 2881 3181 3241 2461 2521 2581 2761 2941 3001 3061 3121 3301 3361 3481 3421

2

25

30

2

S

aaaaccttat aaccaactta tcccctaatd tcccggtaat ggctcgtatg ccatgattac gcttatcgat tccatatcat ttgattattg tatggagttc cccccdccca ccattgacgt gtatcatatg ttatgcccag tgactcacgg ccaaaatcaa cgcctggaga cctccgcggc ggcttggggc tccttcagag gggatctcca gttgttgtat ccgcggtggc catcgctatt taggtgtgggg ctaatccata caaattcaca ggagcttcca gtgccgcaca cgcacggctg ggcagtgtag tttggttctg cggtaggcgt taccgcctat tecttgetee cttgaaaaa gatagcggtt gataataata cactgatgaa tgtcatatga ttatgcttcc acagctatga tcgatatcaa atagcccata cgcccaacga tagggacttt acgtattagt accgatccag gtgacgtaag cggtagcggc gatccaagag atacgttgta catgttgaca ccgcctggca tgttttggca tactgtttt gcttagccta gctcggcagc tegtgttege tggagctcca tacatcaagt cgcaaatggg accgtcagat cttccatta caatactctg ttattattta aacatagcgt caccaccagt aggcagctga aacggtggag gatcaatcat atatttcgtt agattgggct taatagctga cctggctgac ttatattatt gctttacact tggccattgc tcattagttc ggctcttctc ttggaatgaa tcagcaaatt tgaccctgag atacacatct cacacaggaa qaacaaaqc ctgcaggaat acattaccgc gtaacgraa cacttggcag ggtaaatggc aatgggcgtg aatgggagtt gccccattga cgtttagtga ggtgacgata ggctatatgc ggggtcccat gtttttatta gctcatggtc cggtgctgtt cagtacatct agacaccggg cgtgccaaga atgcatgcta tgatggtata caatgcccac atgagcgtgg ccaccagaca gtcaccgtcg tctaaattta atatttttct aagaagatgo ctcatgtcca gttaaggtgg ggcacccag ataacaattt tcactaaagg ateceeeggg aattacgggg tgttcccata gtaaactgcc geggatteee tttggatatt ctatcattgg aacgccattg taattctcgt gattggctat aaatggcccg cgtcaatgac tectacttgg gcagtacatc cattgacgtc taacaactcc aagcagagct cctccataga atgctatagg tatctctatt tttacaggat cgtgcccgca tccggacatg gcctccagcg aggcacagca gtgtctgaaa gcagcggcag 6060606006 cttttctgca tacctatata tttctcaagg actecegttg ctcccctatt atatttgaac gaagtgatgc gaactagtgg gcattggaac tggtaaggta gacggaatgt aaatcattaa tcactcatta ttgtgagcgg caattaaccc tcgagcatca tttatattgg aatagtaatc aacttacggt taatgacgta ccctattga tatgggactt tgcggttttg aggtctatat gctgttttga ggcacaccc acgetteett ttgccacaac actctgtatt gggtacgtgt ttccatgggt gtgtggtgtt cgacgatgaa agtatttacg gtetteacee caaaatgtcg ttattgacca cgccgtcccc ctggtcccat ggccagactt ggtagggtat aagacttaag gtcagaggta actcgttgct cttagaagt attttggtaa aatacattac gtgagttagc gccaagcgcg ggccgctcta aatatqtaca ttgacgtcaa ccctatccaa aagettteet ttgtgtggaa accgctgacc actagttatt cgcgttacat caatgggtgg tcaagtacgc tacatgacct accatggtga cgggactttc cgggaacggt taacagtgga aggccgtggc ggatttccaa gtacggtggg cgccatccac agactctata ttattgacca actgacacgg tatacaacaa catccgagcc ctatacaccc acatggctct cgcgaatctc acgcagatgg tctgataaga tctgagcagt gactgttcct acttcttcgt taatatatt 3601 3661 3781 3841 3901 3961 4141 4201 4261 4321 4381 4021 4081 4441 4501 4561 4621 4681 4741 4801 4861 4921 4981 5041 5101 5221 5461 5581 5161 5281 5341 5401 5521 5641 5701

20

25

30

35

15

gctgcaggag agcctctgga ggggctggag tgcacccgtg ctctgggggc ctcctcagcc ggtgtcgtgg gtcctcagga ccagacctac tgagcccaaa gggggaccg gacccctgag caactggtac tggcaaggag catctccaaa ggatgagctg caggtggcag gcaaatgaat gataacattt gtacaacago cgacatcgcc tcccgtgctg ctacacgcag ctatgagetg ctgctstgga ggcccctgtg ctctggctcc tgaaggtgac cggaggacc ctcctacccg tgaatgtggc cccacctcc agtggagacc gagcctgacg agggagcacc ggccctttt aacactgatg ccagtgccag tatactgtga cccaggtgca ctccagggaa cagactatgo cgttatatct cggggattat tggtcaccgt ccaagagcac aaccggtgac ctqtcctaca gettgggeae gggaggagca tegagaaaac cccatcccg acaagaagt ctgaactcct aggtcaagtt actggctgaa tgatctcccg tctatcccag tggacaagag agaccacgcc tgcacaacca aaaagctttc ccaggatcac aggtgttcag tcataagtga agtcaggcca ctgagagatt aggtggaaga tcactctgtt tcaaggcggg tcacgcatga gggaggaag gcagctacct ctcatcactc ctaaatgggt ttgctgcagg ccccgcacgg tecettagag ccgaaaggta gtccgccagg ggtgggacaa tcaaaaaca tactgtacca gcaccctcct tacttccccg tacaccctgc cagggaaccc acctttccgg ccctccagca accaaggtgg tgcccagcac gadaccrtca gaagaccctg acaaagccgc ctgcaccagg ccagccccca catgaggete gtcaaaggct tccgggatcc agtggggccc tatcataggg aacaactaca aagctcaccq ccagagccga ggacaaacgg taccagcaga geceeetegg agcagcccg tcaccgcgga ctctcgtgcc ctggtgtgtc tacgcggcca agetgeeagg tetaccette ccagcaccc atctcaacat aagagatgat tgcgccagag gaattggggc cttccccctg cctctacagc 900999999 gatgagctgg caaaattgat agccgtatat ggtcaaggac cggcgtgcac ggtgaccgtg taatgccaag cctcaccgtc accacaggtg gcccagcaac atgcccaccg cgtgagccac caaagccctc gacctgcctg ctccgtgatg agtgtcccca aaaacccaag gcagccggag gggtaaagcg tgtttattgg cttgactatc cagcagtggt gcccaaggct tactacttcc caaacgaccc caaggccaca gaaggcagat caaaagctac aagaaataaa cattccttcc caacaacaag tgcagaatgt gccatgactc cagtttcggc gcttggtaaa gaaacgcctg gtattaaaag tcaccatctc ccgaggacac tecetecee gcccatcggt ccctgaccag ttagcaacgt tgggctgcct tgaatcacaa aaactcacac tetteceece tggtggtgga tggaggtgca tggtcagcgt aggtetecaa agccccgaga tcttctcatg actaggtgta acctgtated cagaaaaata actcaactga aggtcagcct agagcaatgg geteettett atgaggacag caatggccac tcctaggtca ttcaagccaa cagtggcctg cagtggccc cctctqccat ccaaacaag ggaagtccca gaaacttcgc ctgccacagc tegttaagea qctttgtcaa aaaggcagat tgggtcggcc gggggagtta tcggggggag ttcactttca agcctgaaag tccaccaagg aactcaggcg atctgcaacg tcttgtgaca tragtetter acagcggccc ctctacttcc gtcacatgcg acgtaccgtg tacaagtgca gccaaagggc accaagaacc gatgcattgc gtggacggcg gtggagtggg gactccgacg acacagccac tactactgtt tctgaggagc caggggaacg aagagcctct gtggtcatct agctcaggga aagctgaccg ggagccgtga accacacct cttgagcagt gtggagaaga gaaggggaa tgcagaggc ttctttgcta ttttcaagag 6001 1909 6181 6241 6301 6361 6421 6481 6541 6601 6661 6721 6781 6841 6901 6961 7261 7741 7021 7081 7141 7201 7321 7381 7441 7501 7561 7621 7681 7801 7861 7921 7981 8041 3161 8101

20

25

30

35

15

tgttaaatca ctccgggatc gatgatgatg aaaacctaaa cgtaaaaaat ttggtgacga ctttgtttgg ccaaadcaaa gtgatttaga acccaattcg aaagaataga aagaacgtgg catgaaggtc ccagccatcc taaaattaaa cttcctdcca caaccatcat ctatcaatgc atcagctttt ctgtttcctg ataaagtgta cgcgcgggga gccaggaacc gagcatcaca tegtgaetge cctgaatggc ttaggggatt tattgagctt aacaccggtt actactatct tcactgcccg ctgcgctcgg ttatccacag taccaggcgt accggatacc cccgttcagc agacacgact gttaaatttt gcaaacagta atgggccaat aggcaaatgc agtggctatt gcacgaaccc tcaatatttt tgtattatct ggggcccggt ttataaatca tggcccacta accaaaatca gaaagcttat cgcggctttt ataacatcat taatccaaat gctgatgact caccadccaa caacccggta ttttacaacg atccccttt agttgcgcag tccactatta atacatgcga tcttctttat gaggggttaa aatgaaaagc gctggattgg atgtataccg atggtcatag agccggaagc tgcgttgcgc aatcggccaa cactgactcg ccagcaaaag ccccctgac actataaaga cctgccgctt tageteaege ggtaatacgg gaaatccagt ccttcccaac taaaattcgc gcaaaatccc aatgatttt tgctatttac tgaagctaaa atttcgcgga gctaaaccaa aaagcacctt gcgatcccac aacgctaaaa tegeceattt cgctttctca tgggctgtgt gtcttgagtc acaggtgttt ctggccgtcg cttgcagcac ggaacaagag atcagggcga attccgtcag gcatatcgca ccctgagctt caagacccgt accacaccgt taaataagtg tggcgtaatc acaacatacg tcacattaat tgcattaatg cttcctcgct actcaaaggc gagcaaaagg ctgttccgac gactcgaggg ataggeteeg acccgacagg gagaattaac acttaatcgc gccaatttat gggtcattat ctgtttactc gtgagctaac atcctgatta gcctcccctc cgcgcgctca caccgatcgc aatattttgt gccgaaatcg gttccagttt aaaaccgtct accttaactt ggctggttat aggttttatt acagtaaaac ataaataaca taatccctgt tgcaggtaaa ttcagatata tgccgttttt tgtaaaagac ctgtaatagc atggcatcgt acaattccac tegtgecage cgctcttccg gtatcagete gegtttttee gtgcgctctc ggtaactatc attgcgcgct aagaacatgt aggtggcgaa ggaagcgtgg cgctccaagc gcgttaccca cagggtcaac gtgaatgatt ttatccgctc tatgccctag agtegtatta gttgagtgtt caaagggcga atgtgtatcc gttatcaaga cacttgtctc ggataataat ggctcaccaa cgagcagtac tgcctaatga ggctactgct ttatcccatt aagaggcccg tgtaagcgtt taaccaatag cttactcaat gataaaaag taaaatagat ctccctgtaa ctaaccaacc gaatactgag atcacgattt taatcaacaa gtgagggtta gggaaacctg gcgtattggg gcggcgagcg cgcgttgctg aagctccctc cgccttatcc taacgcagga tetecetteg gtaggtcgtt ctcaagtcag gctcattttt gctgaaataa cgtaatagcg gaatggaaat caaattggta gcattggtaa caaaggettg gctattcatc atggattaca ttcacccaca ccctatagtg gaaaaccctg ccgagatagg actccaacqt atatgacaag catcagtgct tgcttaaaaa agagettgee gaaagataaa gecetettgg atcgatagca atagacatca acagggaaa gcaataaatc gttcccttta tgtgaaattg ctttccagtc gaggcggttt aatcagggga gtaaaaaggc aataactaag aagcctgggg tegttegget aaaatcgacg ttccccctgg tgtccgcctt tcagttcggt ccgaccgctg 8401 10261 8461 8521 8581 8641 8701 8761 8941 9001 9181 9301 10381 8821 8881 9121 9241 9361 9481 1996 9841 9901 10081 10141 10201 10501 9061 9421 9541 9601 9721 9781 9961 10021 10321 10 5 20 25 30 35

gtaggcgtgt

cgcagcgtga tecttteteg catatcataa gattattgac tggagttccg cccgcccatt attgacgtca atcatatgcc tegetattae actcacgggg gtatttggta tgatccggca acgegeagaa acctagatcc acttggtctg tttcgttcat ttatcagcaa teegeeteea aatagtttgc ggtatggctt ttgtgcaaaa gcagtgttat aaaatcaacq cagtggaacg ttaccatctg gtaagatgct cggcgaccga actttaaaag ccgctgttga tttactttca ggaataaggg agcatttatc aaacaaatag gggactttcc acgttgtatc tgttgacatt cccaacgacc ggtggttacg agcccatata tttcttccct ttttggcacc agcgaggtat tagaaggaca gcagcagatt aaggatcttc atatgagtaa gatctgtcta acgggagggc ggctccagat ttcgccagtt catcaagtgt gcctggcatt gtattagtca tagcggtttg caaatgggcg tiggtagetet gtctgacgct tgcaacttta ctcgtcgttt atcccccatg taagttggcc ttagaaaat catgccatcc atagtgtatg acatagcaga aaggatctta ttcagcatct cgcaaaaag atattattga ctgtagcggc gcattaagcg cggcgggtgt catgtccaac attaccgcca ctcctttcgc gccattgcat tggctgaccg attagttcat cttggcagta taaatggccc tttctacggg ctatctcagc gtacatctac ggattagcag acggctacac gaaaaagagt ttgtttgcaa gattatcaaa tctaaagtat taactacgat cacgeteace gaagtggtcc gagtaagtag tggtgtcacg aacgccaata tgggcgtgga tgggagtttg cccattgacg gagttacatg ttgtcagaag ctcttactgt cattctgaga ataccgcgcc gaaaactctc ccaactgatc ttgaatgtat ggcaaaatgc tcctttttca ID NO:98 (pTnMCS (CHOVep-prepro-HCPro-CPA)) gttaccttcg ttacggggtc atggcccgcc ctagcgcccg cggcatcaga ttggctattg ttcccatagt tcaatgacgg ttgacgtcaa ttggtcatga cggttttggc agtacatcaa tggcctaact cctttgatct tttaaatcaa agtgaggcac gtcgtgtaga ccgcgagacc gccgagcgca cgggaagcta tcaaccaagt atacgggata tcttcggggc actegtgeae aaactgccca ctacttggca acaactccqc actggtaaca acaggcatcg cgatcaaggc cctccgatcg ctgcataatt aaaacaggaa ctcatactct cgaaaagtgc ggatacatat tatattggct cttacggtaa tatttacggt tgccagcgcc tagtaatcaa atgacgtatg cctattgacg ttaagggatt aaaatgaagt taattgttgc tgggactttc ctccaccca aaatgtcgta ggcagcagcc cttgaagtgg gctgaagcca tcaagaagat atgettaate ctgactcccc tqcaatqata agccggaagg tgccattgct tggtgagtac cgctggtagc cggttcccaa ctccttcggt tatggcagca gatgtaaccc cccggcgtca tggaaaacgt tgggtgagca atgttgaata tctcatgage cacatttccc ctgacgcgcc ccacgttcgc tagttattaa atgggtggag ccgctacact tatgtacatt cgttacataa gacgtcaata ggactttcca aagtacgccc tatogcoact ctacagagtt tetgegetet aacaaaccac aaaaaggatc ccatagttgc tttctgtgac catgacetta catggtgatg atttccaagt aaactcacg ttttaaatta acagttacca gccccagtgc gcaacgttgt cattcagctc gttgctcttg taaaccagcc tccaqtctat aagcggttag gatccagttc cgacacggaa cactcatggt tgctcatcat ccagcgtttc agggttattg gggttccgcg 301 181 241 361 421 601 121 10801 10861 5 481 541 199 10621 10681 10741 10981 11041 11101 11341 11461 11641 10921 11161 11221 11281 11521 11581 11401 11701 11761 11821 11881 11941 SEO 2 20 25 30 35

ccgcctatag aatccataac cttcagagac aattcacata gatctccacg agettecaca cttgctccta gccgcacaag cacggctgac tgttgtattc cagtgtagtc gactaacaga actcgatatt caacagetta acttggccgt ctttatctgt gcaaaaacga ttatgagaaa agccgacctt ctttaaagtg aagaggaaaa acatgatatg tccaatctca ctcgacacgg tatctattcg cggactaggc aatcgccctg ttgggacaag cttaggcatg ggctgcaacc aggggatcgc tttatcatca ttatgattga attgttaggt gccatgggtt cttggggcct cgtcagatcg cgatccagcc gacgtaagta ctgtttttgg atactctgtc attatttaca catagcgtgg teggeagete cggtggaggg atagctgaca ccatgtgcga taaaacgact ctcttaccga gaatcgaccg ggcatgctag gtgatgctgg cggcaaagga aacttgttaa ttagcctata ttccattact attgggctcg atattcgtga ctgttactct ctaaaaqcaa aaaatcagcg gtcctgccta agaaacaagg ggaaattatg cagcaattaa gtagcggcgg ccaccagtgt gcagctgagt accaatttct taagtcgagt tcagcaactt tcatgctgct caacagttcg acttactcgt aaaggtgcct tttagtgaac gttcatgctc ctgttataag acaccaggac gcatgctata atggtatage tgacgatact ctatatgcca ggtcccattt ttttattaaa ctcttctccg tcatggtcgc atgcccacca gagcgtggag gaaga+gcag gtgccgttaa accagacata caccgtcgga aaaactctca aacatcaaac gtataccgtt gactggtctg cacggtcgtt aaagctcatg ctcattgtca tggaaaccta aagaggctga aaaggccgaa tactcagcgt acacccaaac gacttgaaaa cgttttgata acaagggaag tacgetttgg gttggtgacc tgccaagagt gaattacact tggtactggt aacqtactct agcggaaaac tttaggctat gcagageteg tccatagaag ggattccccg tggctcttat gctataggtg cggacatggg gcacagcaca agcggcagaa ეენენენენე tttctgcagt attetgeece gcgactcgct tgtacttgtt agtcgcacta atgttcaaag gaagctgggt acctaaaatc tgaaattcga aaccttccga cagctcagag cagaaatcga cacacatggt cccctattgg tctctattgg tacaggatgg ctccagcggc gtctgaaaat tecegttgeg acttgactgt atctcgctct gcttgcgggc ctacacaata cgggaaaagc tactcagtgc tgcccgcagt aacaaaacat caccacaccg tgcgagette gtctatataa tgtttgacc gcttccttat tctgtatttt gtacgtgttc ggtcccatgc gacttaaggc cagaggtaac tegttgetge gccacgcatt tccacaaaga cgatgcccat tttcagagca aatccgttga actcaaagac tattgtataa accacccgtc ggcattctgg cacacccctt attgaccact gccacaacta ccgtcccccg ccagacttag tagggtatgt ccatgggtct ctctttacca ccaaagcgag taccgagtaa acttacctgt agattgaaga gccgaacgag ctaacacagt teegggatet attggaacgc cagacctagg aaaatttatt aaaaaaat taacatgttg gegtteeege gcgagcattc actcattgtc ccatccacdc ggaacggtgc atggctcttt tgacacggac tgagcagtac ctgttccttt ttacacgact acgttggctt aacctdccaa aatcgtcacc tegggcaata ccatggtata gtacaatatg tgccaaattc ctagcaacta aagcgaatgc atgetteaac gaagttttgc ctactagctc ctttaaaaat acggtgggag actctatagg atacacccc attgaccatt tacaacaacg cgaatctcgg gccgtggcgg gcagatggaa cttatggtat ctacqccata cacttccagg tctagagcga acagtggagg tcatctagtc tccgagccct tgataagagt 1741 2341 901 1021 1081 1321 1441 1501 1561 1621 1681 1801 1861 1921 1981 2041 2101 2161 2221 2281 2401 2521 2581 2641 2761 2821 2881 1141 1201 1261 1381 2461 2701

15

2

25

39

35

attactggta ggtaaaaatc tggaattgtg ttagctcact gcgcgcaatt ctctagaact tgacctcgag cctgatacct agaaaccatc aaggcaattt cttcctaatd ttttagtgat tagggacaga aagattgttg ttcttagtgg tcagttgata tctgtgacct gataaaaagt ctttcttgg agataacatt cagctgtgct tcttaagtcc aaaggaagga aagtaattct gtacttttt ttgctattgt aattcctctg caaqttcata ttttgctgta atttaaatat gctccttcac gatcaaatga atcccattat tactgtttg tctgccttag cttatcccta ccttqaatac acttaaagct taatgatttt gtaatgtgag gtatgttgtg attacgccaa 959999959 tegatacege aggagettga cagctcagag tegacattea ccctcatgaa taatgtgctc tctgagtagg cttagctgac tatccttttg actgttggtt ctgctttgca caactgcaag tcataggtag tttcagact acaaaaaaa taccatcttg tggtatagga gaattatgtt atggtttctt tgaaatatat rgttttcaca tgaatttcac aggaacactg tggagctatg aaaaagatag qtctcagttg ataggtcacg atacaagaag tttcttaaag aaatggttgg atgaatcccc tatgaccatg taataaaac aaaaaattag aagagaacca tatgatcccg cttccggctc atcaagctta tcccacaaa ctccaccgcg tcacatccaa ggcaatccat ttgatiattt agagaacaat aatctcaagt tattcatgag ggtaaccttc gcaggcaaga ctggacttca actatttgta gacatgcatt acaaaatgaa ttatttgcac atgtcctgaa cttctataac gcagcagagc cttcagctag agaggtcaga ctcctccaqc acagtatata aagatctttg ttgcattgta tcacacccat cttcactcta caaqaaaata cacgttgtag tagtaataag attttttctgt ctgatttaac ttattgataa atgaacttga ctgagcactg acactttatg gactatgaac aaattgatcc catcttqtca aggaaacagc aaagctggag ggaattcgat aacaacacaa aaagacaaac gtgagctcac ggtatgcaat ccatgaatgg aggagaacaa teegagtett ataggaaagg gtgtgacaaa acctatccca tctqcaqact atcacaaaaa gtaattaaaa caaaaaata gtagatttcc actgactaaa atctctatgt agtgctttac aggttcccat ttattcatgg tataaagcat agaacaatag accatgaaca gagtattggt tgcgtctgct cctattttgt ggctcagatg cctacaacca acaacaaaa cttgattata attggttgga cattgtcagc ctcgttgacc ggtggataca cccaggcttt aatttcacac aaagggaaca aaacagggga ccgggctgca atgccaggtg aatatgcgac atttctctct gctttgacct cttcaggtc aatgcctggc ctagaaagct acagcaccaa agtggcaaag atccattgcc cacatctatc agcaaaagac agacatccag gattgcagtg taaatcaaac atttcaatac gtcaaacaat caaggagaat agaaaatggc tgtccctcga ggccattaag aaaagtttgt ggaaagaaag tggtctgtta tccatctcta atttgaaatt ctttatttct tgcctagcat tgaacattat agcggataac gattttcttc gatgcctatc aggtaaacgc aatgttaatt attaagttaa cattaggcac aaccctcact agtggatccc ctgcagaaa actgatggct gagcaaaatg ccacactcac tcaaaattgt cccacaatga tggataagag tactgggaag tcatctgcca gcttttataa gacaagcatt ctgaaataga gagaaacaaa gcttaacaga tcctcagaaa aacagacttc aattatgttt gggcagggcg cttactctct cctgtggtgt tcagacttag ccccatttt caattctatt atattatgat tcataaacac tccagctata atgcatgctt acttatgtcc 3301 3421 3361 3541 3481 3601 3661 3721 3781 3841 3901 3961 4021 4081 4141 4201 4261 4321 4381 4441 4501 1961 1621 1681 4741 4921 **4981** 5041 5221 4801 1861 5101 5161 5281

109

25

30

WO 2004/067743

aattaaaaac attaaaattt ttggttctgg gcctctggat gggctggagt gcacccgtga caaatgaata ataacatttg tctgggggca gtgtcgtgga tecteaggae cagacctaca gagcccaaat ggggaccgt accctgagg aactggtacg tacaacagca ggcaaygagt atctccaaag gatgagctga gacatcgccg cccgtgctgg aggtggcagc tacacgcaga gagetgacae cctcctctq gtgggtgggt ctgcaggagt tcctcagcct tctggagatg ggctccagct ggtgactact gagaccacca cctgtggtgg cccagtcttc tgttacccag gttatatctg ggggattatg accggtgacg ggtcaagttc aactaaaatc atcaatcatt tatttcgttt cgtgttcgct ccaggtgcag ctcctgtgca tccagggaag ggtcaccgtc caagagcacc tatcctacag cttgggcacc tgaactcctg ggaggagcag cccatcccgg gcacaaccac caagaaagtt cgagaaaacc ctatcccago ggacaagagc gctttcctat gatcacctgc gttcagcgga ggcgggagtg aagggaacct agactatgct gateteeegg ctggctgaat aggccaggcc ggaagatgaa tetattecca gaccacgcct gagattetet aagtgactcc aqcaatattt gaaaacacat gacctattga cagccagtgg ctaaatttaa tatttttctt cgaaaggtac cccttagagt tacgacagga gtgggacaac caaaaacac actgtaccac caccetecte acttccccga cctttccggc tgtgtctcat caaacttctg agggancet ccaaggtgga gcccagcacc aagaccctga caaagccgcg tgcaccagga ggatccctga cctcggtcac gccccgtcaa cctccagcag acaccctcat cagcccccat acaccctgcc tcaaaggctt acaactacaa cagagccgaa aaacggccag agcagaagtc atagggaggt agctcaccgt atgaggetet gggcccaggt atggaacatg atccaaaatg ttctcaagga atgagetggg ttccccctgg gtgaccgtgc agaacaggca ccccagggct acctatatat gcgccagagc aaaattgatg agagatgatt aattggggcc teegtgatge tccccaggac tgtcaaaggt ccgggggggt gccgtatatt gtcaaggact ggcgtgcaca cccagcaaca tgcccaccgt aaacccaagg gtgagccacg aatgccaaga ctcaccgtcc aaagccctcc ccacaggtgt cagccggaga ctctacagca ggtaaagcgc tattggtacc actatcagtg agtggttatc aaggctgccc gcagatagca acctgcctgg cgaccctccg gccacactgg ctgaagctca tctccattca tgtggtgtt gacgatgaat ggtggtggac ggtctccaac atggattagc atttctatgg gctatatatt agtttcggct cttggtaaag caccatctca cccatcggtc gggctgcctg ggaggtgcat ggtcagcgtc aaacgcctgg tattaaaagc cgaggacaca tagcaacgtg aactcacaca cttcccccca gccccgagaa ggccaccttg ccctcccccg cctgaccagc gaatcacaag ggtcagcctg gagcaatggg ctccttcttc cttctcatdc ggtgtcagtg aaaatatgtt ggacagcaaa aactgacagc aggtcagccc agccaacaag ggcctggaag cctgtctccg taatatttgc aacagtcctg ccattaaatg cacaattcag cttcttcgtg ctttgtcaac cggggggagg tcactttcag gcctgaaagc ggggagttat ccaccaaggg actcaggcgc cttgtgacaa aggagettea aatatattc gggtcggccg aaggcagatt caqtcttcct cgtaccgtgt acaagtgcaa ccaaagggca tggagtggga actccgacgg agagcetete agccaccctc cattgccaga tcatctatga actettactc cagoggooot tctacttcct tctgcaacgt tcacatgcgt tggacggcgt ccaagaacca aggggaacgt cagggacaat tgaccgtcct ccgtgacagt 5581 5641 6301 6541 6601 6781 6901 7561 5821 5881 6001 6121 6181 6241 6361 6421 6481 6661 6721 6841 1969 7021 7081 7141 7201 7261 7321 7381 7441 7501 7621 7681 5941 6061 2 15 2 30 35 25

ctgatgtgca tgagatttt gcgtaatagc gactccaacq ctgacgette agcaccgtgc ctttttgaag tgtggcttct tgccaggcac agcaaaatgg tttcgcttca gccctatagt ggaaaaccct cgaatggaaa agctcatttt accgagatag tcatcagtgc gtgcttaaaa catatgacaa tgaaagataa catcgatage ccaaattggt aagagcttgc tgccctcttg aaataactaa tgcattggta gatagacatc aacagggaaa tgcaataaat acaaaggett cactgaaata tgctattcat aaagcctggg tgttcccttt gtgtgaaatt ctacctgage ttatctgtgc taccaaaatc attaggggat gcatgaaggg gggaaggcc gcacggaaca tcactctgaa atgggtttcc tgcaggccag gtcgtgactg tattttccaa tacccaattc tegecagetg gcctgaatgg ttgttaaatc aaaagaatag actccgggat tgatgatgat aaagaacgtg aaaaacctaa ttattgagct tcgtaaaaa tocagocato ataaaattaa tetteetgee tttggtgacg cactactatc acatgaaggt taacaccggt tettette ccaaccatca gctatcaatg gatcagcttt cataaagtgt acgcgcgggg gctgcgctcg getgttteet gttatccaca cggccagcag gccaggtcac caacatctaa caccccttgc tccagttcaa gtgttttgta 8999966699 gttttacaac cttacaaatc cgcggaggga cgtgccccc ccettcctcg cagttgcgca gtccactatt ggaaagctta catccccctt cgttarattt atggcccact aatacatgcg ccgcggcttt atcttcttta aataacatca tgaggggtta aatgggccaa tgctgatgac ataatccaaa ccaccagcca tagtggctat taatgaaag aaggcaaatg tgctggattg ttgcgttgcg gaatcggcca gatgtatacc catggtcata gagccggaag tcactgactc cggtaatacg acaaagcaac aacaagtacg qaatqttcac tgactcctct getteetetg aataaaatct cetteceeag atgagagaaa tgattaacag agctacagct cccctcgagg actggccgtc ccttgcagca cccttcccaa ggcaaaatcc tgcatatcgc ttaaaattcg tggaacaaga taatgattt tatcagggcg cattccgtca ttgctattta ttgaagctaa tatttcgcgg cgctaaacca agcaaacagt agcgatccca aaacgctaaa ccctgagct taaagcacct ttaaataagt ttggcgtaat ctcacattaa ttcgcccatt caccacaccq cacaacatac gtcgtgccag ctgcattaat getteetege cactcaaagg ccaagacccg gtcccacaaa tgccattgct ggccctgca cttcgcgcca cacagcaaga taagcacatt actgctgccc ccctagatcc ccattgcct acgegegete aacttaatcg ggccgaaatc tgttccagtt aaaaaccgtc cgagaattaa tggctggtta ggccaattta taggttttat ataatccctg gcaccgatcg taatattttg caccttaact agggtcatta cctgtttact tacagtaaaa agtgagctaa tataaataac atgcaggtaa cttcagatat ctgccgtttt gtgtaaaaga tctgtaatag aatggcatcg aattgcgcgc cacaattcca gegetettee ggtatcagct gagggccctc ttgctactgc ggcgttaccc agcagtggaa agaagacagt ggggaggaaa caagagtcgt caacttggct attacatatg cccacattat gagtcgtatt gaagaggccc ttgtaagcgt ttaaccaata ggttgagtgt tcaaagggcg gatgtgtatc tcagggtcaa acttactcaa ataaaataga cqataaaaa ggttatcaag aggataataa agtgaatgat aggeteacea actecetgta actaaccaac agtgagggtt gtgcctaatg gcacttgtct ccgagcagta ggaatactga catcacgatt ataatcaaca gttatccgct cgggaaacct tgcgtattgg tgcggcgagc 7981 8101 8161 8281 8341 8221 8641 8401 8461 8521 8581 8701 8761 8821 8881 8941 900I 9061 9121 9181 9241 9301 9361 9421 9841 9661

25

30

WO 2004/067743 PCT/US2003/041269

	10141	. ataacgcagg	aaagaacatg	tgagcaaaag	gccagcaaaa	ggccaggaac	cgtaaaaagg	
	10201		ggcgttttc	cataggetee	gccccctga	cgagcatcac	aaaaatcgac	
	10261			aacccgacag	gactataaag	ataccaggcg	tttccccctg	
	10321			cctgttccga	ccctgccgct	taccggatac	ctgtccgcct	
	10381		gggaagcgtg	gegetttete	atageteaeg	ctgtaggtat	ctcagttcgg	
	10441	. tgtaggtcgt	tegetecaag	ctgggctgtg	tgcacgaacc	ccccgttcag	cccgaccgct	
	10501	-	_	cgtcttgagt	ccaacccggt	aagacacgac	ttatcgccac	
	10561			aggattagca	gagcgaggta	tgtaggcggt	gctacagagt	
	10621		gtggcctaac	tacggctaca	ctagaaggac	agtatttggt	atctgcgctc	
	10681	-	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	aaacaaacca	
	10741	_	cggtggtttt	tttgtttgca	agcagcagat	tacgcgcaga	aaaaaaggat	
	10801	_	tcctttgatc	tttctacgg	ggtctgacgc	tcagtggaac	gaaaactcac	
	10861	_	tttggtcatg	agattatcaa	aaaggatctt	cacctagatc	cttttaaatt	
	10921		ttttaaatca	atctaaagta	tatatgagta	aacttggtct	gacagttacc	
	10981		cagtgaggca	cctatctcag	cgatctgtct	atttcgttca	tccatagttg	
	11041	. cctgactccc	cgtcgtgtag	ataactacga	tacgggaggg	cttaccatct	ggccccagtg	
	11101	. ctgcaatgat	accgcgagac	ccacgeteac	cggctccaga	tttatcagca	ataaaccagc	
	11161		ggccgagcgc	agaagtggtc	ctgcaacttt	atecgeetee	atccagtcta	
	11221	ttaattgttg	ccgggaagct	agagtaagta	gttcgccagt	taatagtttg	cgcaacgttg	
	11281	. ttgccattgc	tacaggcatc	gtggtgtcac	gctcgtcgtt	tggtatggct	tcattcagct	
	11341	ccggttccca	acgatcaagg	cgagttacat	gateceecat	gttgtgcaaa	aaagcggtta	
	11401	geteettegg	tectecgate	gttgtcagaa	gtaagttggc	cgcagtgtta	tcactcatgg	
	11461	ttatggcagc	actgcataat	tctcttactg	tcatgccatc	cgtaagatgc	ttttctgtga	
	11521	_	ctcaaccaag	tcattctgag	aatagtgtat	gcggcgaccg	agttgetett	
	11581	gcccggcgtc	aatacgggat	aataccgcgc	cacatagcag	aactttaaaa	gtgctcatca	
	11641		ttcttcgggg	cgaaaactct	caaggatctt	accgctgttg	agatccagtt	
	11701	cgatgtaacc	cactcgtgca	cccaactgat	cttcagcatc	ttttactttc	accagcgttt	
	11761	ctgggtgagc	aaaaacagga	aggcaaaatg	ccgcaaaaa	gggaataagg	gcgacacgga	
	11821		actcatactc	tteettttte	aatattattg	aagcatttat	cagggttatt	
	11881			tttgaatgta	tttagaaaa	taaacaaata	ggggttccgc	
	11941	gcacatttcc	ccgaaaagtg	ccac				
>								
SEQ		ID NO:99 (pTnMCS(CHOvep-prepro-HCPro-Lys-CPA))	CHOvep-prepr	:o-HCPro-Lye				
	בל לט	1 ctgacgcgcc ctgtagcggc	ຫຼ	gcattaagcg cg c ctadcdcccd c	eggegggtgt ggtggttaeg e g eteetttege tttetteeet		cgcagcgtga E tcctttctcq	
	_			ttggctattg g	gccattgcat a		catatcataa	
	181	tatgtacatt t tagttattaa t	tatattggot o taotaatcaa t	catgiccaac a ttacqqqqtc a	attactcat u attacttcat a	egtegacate y agcccatata t	garrarrgar tggagttccg	
							1	

attgacgtca togotattac aaaatcaacq atcatatdcc actcacgggg cctggagacg teegeggeeg ccgcctatag ggtgtgggtt aatccataac cttcagagac aattcacata gatctccacg agettecaca cttgctccta cacggctgac tgttgtattc gccgcacaag cagtgtagtc gactaacaga actcgutatt cttggggcct caacagctta gtaggcgtgi acttggccgt attgttaggt gcaaaaacga ttatgagaaa agccgacctt ctttaaagtg acatgatatg tatctattcg ctttatctgt aagaggaaaa tccaatctca gggactttcc ttttggcacc catcaagtgt gcctggcatt gtattagtca tagcggtttg caaatgggcg cgatccagcc gacgtaagta ctgtttttgg ttagcctata atactctgtc attatttaca catagogtgg gtagcggcgg teggeagete cgtcagatcg ttccattact attgggctcg ccaccagtgt gcagctgagt cggtggaggg atagctgaca ccatgtgcga atattcgtga taaaacgact ctcttaccga gaatcgaccg ggcatgctag ctgttactct accaatttct gtgatgctgg tcagcaactt ctaaaagcaa cggcaaagga aacttgttaa taagtcgagt aaaatcagcg cttggcagta taaatggccc tggctgaccg aacqccaata gtacatctac tgggagtttg cccattgacg atggtatagc tgggcgtgga tttagtgaac acaccgggac ctcttctccg tgccaagagt gcatgctata tgacgatact ctatacgcca ggtcccattt tttattaaa tcatggtcgc atgcccacca gagcgtggag gtgctgttaa gaagatgcag accagacata caccgtcgga aacatcaaac gaattacact aaaactctca gtataccgtt gactggtctg aaagctcatg tggtactggt tggaaaccta cacggtcgtt ctcattgtca aagaggctga aaaggccgaa acacccaaac tactcagcgt atggcccgcc tcaatgacgg ttgacgtcaa ggattccccg tggctcttat ttcccatagt aaactgccca ctacttggca agtacatcaa acaactccgc gcagageteg gctataggtg ccctattgg cggacatggg tecegttgeg tttctgcagt attetgeece aacaaaacat tccatagaag tctctattgg tacaggatgg tgcccgcagt ctccagcggc gcacagcaca gtctgaaaat cgcgcgcgcc gcgactcgct agcggaaac agcggcagaa acttgactgt agtcgcacta acctaaaatc tgaaattcga tgtacttgtt atgttcaaag caccacaccg tttaggctat atctcgctct gaagctgggt cttacggtaa atgacgtatg cctattgacg tgggactttc cggttttggc attggaacgc tatttacggt ctccaccca aaatgtcgta gtctatataa tgttttgacc cacacccctt gcttccttat attgaccact gccacaacta tctgtattt gacttaaggc gtacgtgttc cagaggtaac tagttgatga ctctttacca ccgtcccccg ggtcccatgc ccagacttag tagggtatgt ccaaagcgag ccatgggtct gccacgcatt tccacaaaga cgatgcccat tgcgagette tttcagagca taccgagtaa cagacctagg actcaaagac tattgtataa accaccogto aatccgttga acttacctgt cqttacataa gacgtcaata aagtacgccc atgggtggag catgacctta catggtgatg atttccaagt ggactttcca acggtgggag ggaacggtgc actctatagg atacacccc attgaccatt atggctcttt tgacacggac cgaatctcgg ccatccacgo tacaacaacg acagtggagg gcagatggaa tgagcagtac gccgtggcgg tgataagagt ctattccttt ttacacgact acgttggctt aacctdccaa gegtteeege aatcgtcacc tegggcaata cttatggtat gcgagcattc tcatctagtc taccaaattc teegageeet ccatggtata gtacaatatg acteattgte ctagcaacta 721 781 841 1021 196 1081 1141 1201 1261 1321 1381 1441 1681 1501 1621 1741 2341 1561 1801 1861 1921 1981 2041 2101 2161 2281 2401 2461 2221 2521

25

30

35

WO 2004/067743 PCT/US2003/041269

tggaattgtg aatcgccctg attactggta ggtaaaaatc gcgcgcaatt tgacctcgag cttcctaatd ttgggacaag ggctgcaacc aggggatcgc tttatcatca ttatgattga aaagtatatt ttcctgacgg agaaaccatc tctgtgacct ttttagtgat gataaaagt tagggacaga ctttcttgg aagattgttg agataacatt ttcttagtgg tcagttgata tcttaagtcc aaaggaagga aattcctctg cttaggcatg tccaaqaaqt ttagctcact ctctagaact cctgatacct aaggcaattt cagctgtgct aagtaattct tactgtttgt gtcctgccta ggaaattatg tctgccttag acttaaagct gtaatgtgag gtatgttgtg tcacatccaa aggagettga tatccttttg cttagctgac tgttttcaca tgaatttcac tcatgctgct cagcaattaa ccttgaatac gtggcggccg cagctcagag taatgtgctc tctgagtagg caactgcaag tcataggtag gaattatgtt tgaaatatat agaaacaagg caacagttcg acttactcgt aaaggtgcct cttatcccta taatgattt attacgccaa tcgataccgc tcgacattca ccctcatgaa actgttggtt ctgctttgca acaaaaaaa taccatcttg tggtatagga ttttcagact gacttgaaaa acaagggaag tacgetttgg gttggtgacc aagagaacca atgaatccc ttgcttcttt agagaacaat aatctcaagt actatttgta ggtaaccttc gacatgcatt gcaggcaaga ctcctccagc cgttttgata gttcatgctc aacgtactct ctgttataag aaatggttgg taataaaac aaaaaattag tatgatrccg cttccggctc tatgaccatg ctccaccgcg atcaagctta tcccacaaa ggcaatccat ctggacttca tattcatgag gcagcagagc acaaaatgaa ttatttqcac acaqtatata cttcagctag atgtcctgaa agaggtcaga cttctataac ctgatttaac aggaaacagc ggaattcgat gactatgaac aggagaacaa gtgagctcac gtgtgacaaa aaccttccga gcttgcgggc cagaaatcga ctacacaata cacacatggt cgggaaaagc tactcagtgc ttattgataa atgaacttga aaattgatcc ctgagcactg catcttgtca acactttatg aaagctggag aacaacacaa ccatgaatgg aaagacaaac tccgagtctt ataggaaagg acctatccca atcacaaaaa gtaattaaaa caaaaaaata gtagatttcc actgactaaa atctctatgt agtgctttac agaacaatag accatgaaca cagctcagag ggtatgcaat tctgcagact aaagggaaca aaacagggga aatatgcgac agattgaaga ggcattctgg cttgattata attggttgga cattgtcagc ggtggataca cccaggcttt aatttcacac ccgggctgca atgccaggtg agtggcaaag gctttgacct ctttcaggtc aatgcctggc gtcaaacaat cacatctatc gattgcagtg taaatcaaac tgtccctcga taacatgttg ctaacacagt aaaatttatt tccgggatct aaaaaacaat acaacaaaa ctcgttgacc acagcaccaa atttctctct atccattgcc ctagaaagct agcaaaagac agacatccag caaggagaat agaaaatggc atttcaatac gccgaacgag gattttcttc aaqcqaatqc cacttccagg ctactagete tgcctacatc gatgcctatc aatgttaatt cattaggcac agcggataac gagcaaaatg ccacactcac tggataagag gagaaacaaa atattatgat ctacgccata atgetteaac tctagagcga ctttaaaaat tgaacattat aggtaaacgc attaagttaa agtggatccc ctgcagaaaa gcttttataa cccacaatga 999cagggcg tactgggaag ctgaaataga gcttaacaga gaagttttgc aacceteact actgatggct tcaaaattgt gacaagcatt cttactctct cctgtggtgt tcagacttag ccccatttt 3361 3541 3661 2881 2941 3001 3061 3121 3181 3301 3421 3481 3721 3781 3841 3901 3961 4021 4081 4141 4261 4501 4621 2761 3241 3601 4321 4381 4441 4561 4681 4741 4801 2 15 25 30 20 35

tcatctqcca

tettgttete caagttcata ttttgctgta atttaaatat geteetteac gatcaaatga aataattgct atcccattat tctcccatcc aattaaaaac taacccaatc catctcgtga attaaaattt ttggttctgg ctgcaggagt gcctctggat gggctggagt gcacccgtga caaatgaata ataacatttg tecteageet tctgggggca gtgtcgtgga tectenggae gtgggtgggt cagacctaca gagcccaaat ggggaccgt tacaacagca ggcaaggagt atctccaaaq tatgagctga acccctgagg aactggtacg gatgagctga gacategeed cccgtgctgg aggaacactg tggagctatg aaaaagatag gtctcagttg ataggtcacg atacaagaag cacactaact tttcttaaag cccagtcttc tigttacccag aactaaaatc aagggaacct atcaatcatt tatttcqttt cgtgttcgct ccaggtgcag ctcctgtgca tccagggaag gttatatctg agactatgct ggggattatg caagagcacc ggtcaccgtc accggtgacg cttgggcacc ggtcaagttc tgtcctacag caagaaagtt tgaactcctg gatctcccgg ggaggagcag cgagaaaacc cccatcccgg ctatcccagc gcacaaccac aaagctttcc ctggctgaat gaccacgcct ggacaagagc aagatctttg ttgcattgta tcacacccat atttttctgt gacctattga cttcactcta caagaaaata cacgttgtag tagtaataag agcaatattt caaacttctg cagccagtgg cgaaaggtac gtgggacaac gaaaacacat ctaaatttaa tatttittt cccttagagt tccgccaggc caaaaacac actgtaccac acttccccga cctttccggc agggaaccct caccttcttc ccaaggtgga gcccagcacc tgcaccagga cctccagcag acaccctcat aagaccctga caaagccgcg cagccccat acaccctgcc tcaaaggett acaactacaa cagagccgaa agctcaccgt atgaggetet ttattcatgg cctattttgt aggttcccat ccccagggct ttctcaagga tataaagcat tgcgtctgct ggctcagatg cctacaacca atggaacatg agaacaggca atccaaaatg tgtcaaaggt acctatatat gegecagage atgagctggg aaaattgatg aattggggcc gagtattggt cc39393999t gccgtatatt ttccccctgg agagatgatt gtcaaggact ggcgtgcaca gtgaccgtgc tccgtgatgc cccagcaaca tgcccaccgt aaacccaagg gtgagccacg aatgccaaga ctcaccgtcc aaagccctcc cagccggaga ctctacagca ggtaaagcgc ccacaggtgt acctgcctgg ggccattaag atttgaaatt aaaagtttgt ggaaagaaag cttatttct tgcctagcat tggtctgtta tccatctcta ctgaagctca atggattagc tctccattca atttctatgg tgtggtgttt gacgatgaat agtttcggct tattaaaagc gctatattt cttggtaag caccatctca cgaggacaca cccatcggtc ggtggtggac ggtcagcgtc aaacgcctgg gggctgcctg cctgaccagc ccetcccccg cttccccca ggtctccaac tagcaacgtg gaatcacaaq aactcacaca ggaggtgcat gccccgagaa ggtcagcctg cttctcatgo gagcaatggg atcattatta cctgtctccg tcctcaqaaa tccagctata aacagacttc ctggttgtaa aacagtcctg tcataaacac acttatgtcc aattatgttt taatatttgc atgcatgctt ccattaaatg cacaattcag aatatattc ctttgtcaac tcactttcag cttcttcgtg C999999999 gggtcggccg ccaccaaggg actcaggcgc aaggcagatt gcctgaaagc tctgcaacgt cttgtgacaa ggggagttat tctacttcct tcacatgcgt tggacggcgt cagoggcoot cagtcttcct cgtaccgtgt acaagtgcaa ccaaagggca ccaagaacca tggagtggga acteegaegg agagcetete aggggaacgt 5041 5101 5161 5221 5341 5401 5521 5881 5461 5581 5641 5701 5761 5821 5941 6001 6121 6061 6181 6241 6421 6301 6361 6481 6541 6601 6661 6721 6781 6841 6901 6961 7021 7081 7141 7261 7201

20

25

30

35

2

gccctttttg ctgggaaaac tggcgaatgg gtggactcca gattcatcag tgctctggag gccctgtgg ggagggacca gtggagacca gggagcaccg tcctgagatt cagtgccagg caaagcaaaa tgettteget ttegaaatat ctggcgtaat atcagctcat tagaccgaga gatcatatga gatgtgctta gcttgaaaga acgaaataac tctggctcca ccaccctcct tcctacccgg agcctgacgo gaatgtggct taaaagagct aaatgccctc ggtcatcgat atcccaaatt ggttgcattg ttggatagac taaaacaqqq tcatgctatt gaaggtgact acactgatgt ggtgttcagc caatatttc gtattatctg tttttgttaa caggatcacc tgagagattc ggtggaagat cactctgttc cataagtgac caaggcggga tgctgcaggc aacgtcgtga ctttcgccag attaaagaac actactccgg atcattaggg ttatgatgat tcatttggtg ttaacatgaa gtcaggccag cagctacctg cacgcatgaa ggagggaagg cccgcacgga tegteactet taaatgggtt cggtacccaa gcagcctgaa atcaaaagaa gcgaaaaacc tttttattga ttatcgtaaa aaatccagcc aagccaacca caataacacc gactctttgt ccaataaat atgcactact ccgggatccc gcagccccgt tetegtgeee ctgcccttcc caggtgtttt aggggggcc gtcgttttac gcacatcccc tegegttaaa gcgatggccc tcaggaaagc atttagtggc ccgtaatgaa accagcagaa gtggggccca atcataggga cccctcggt tggtgtgtct acgcggccag caccgcggag totcaanato aaatccagtt caacagttgc tcccttataa agagtccact ttttaccaaa cgcaatacat ttaccgcggc taaatcttct cggaataaca gcttgagggg ccaataatcc agtaatgggc ccttgctgat aaaaggcaa gctgccaggt cagcauccct ccaccag tatttgaagc gtgtccccag gtttattggt ttgactatca agcagtggtt cccaaggctg aaggccacac aaggcagata aacaacaagt aaaagctaca gcagaatgtt ccatgactcc gctgcttcct agaaataaaa cccatgagag tcctgattaa cctcccctcg ctcactggcc tegeettgea tegeeettee ttgttaaaat atcggcaaaa gtttggaaca gtctatcagg acttaatgat taacattccg ttattgctat aaacgctaaa aacagcaaac ctgtaaagca agaccaagac aaacgaccct atteetteee ttatqcatat ttatatttcg taaagcgatc tataaacgct actcccctga ctcggtgtca aatggccacc ctcaactgac cctaggtcag tcaaqccaac caaacaagc gaagtcccac agtggccct cgttaagcac gctactgctg agataggttt taatacagta agaaaaatat tgaggacagc agtggcctgg aaacttcgcg ctctgccatt tgccacagca atgccctaga tatcccattq cccaacttaa ataggccgaa tgttgttcca gcgaaaaacc atccacctta caacgagaat caatggctgg aagagggtca tetectgttt gattataaat ccaataatcc gtactgccgt tgagtgtaaa attacgcgcg cccgcaccga cgttaatatt aaaggccaat gtaatgcagg aaccttcaga cacagccacc tggtcatcta gctcagggac gagccgtgac cctggcgtta atgcattgcc actactgtta agctgaccgt ctgaggagct ccacaccctc ttgagcagtg tggagaagac aagggggagg gcagaggcc tctttqctac caccaacttg tgctcagggt ttgggttatc ggtagtgaat tttcaagagt tggattacat tcacccacat agtgagtcgt agcgaagagg aaattgtaag tttttaacca tagggttgag acgtcaaagg caagatgtgt aaacttact tgccgataaa taaataaat taagcacttg agcaggataa qtaaqqctca atcactccct aaactaacc aatccgagca cttggaatac 7441 7501 8341 7381 7561 7621 7681 7741 7801 7861 7921 8041 8101 8161 8221 8281 8401 8461 8521 8581 8641 8701 8761 8821 8881 8941 9001 7981 9061 9121 9181 9241 9301 9361 9421 9481

2

25

30

35

15

atgcgctgaa ttttgttccc cctgtgtgaa teggtegtte tgtaaagcct cccgctttcc gggagaggcg acagaatcag aaccgtaaaa cacaaaaatc gcgtttccc tacctgtccg tatctcagtt cagcccgacc gacttatcgc ggtgctacag ggtatctgcg ggcaaacaaa agaaaaaag aacgaaaact atccttttaa tetgacagtt tctggccca tcatccatag ttgcgcaacg gcaataaacc tccatccagt gcttcattca aaaaagcgg ttatcactca tgattttatg ataggggttc aaagtgctca ttqaqatcca ttcaccagcg agggcgacac tatcagggtt ccgagttgct ttggctatca acggttatcc accgatcago atagctgttt aagcataaag gegeteactg ccaacgcgcg ctcgctgcgc aaaggccagg tgacgagcat aagataccag gcttaccgga acgetgtagg accccccgtt ggtaagacac gtatgtaggc gacagtattt gattacgcgc cgctcagtgg gggcttacca ctcttgatcc cttcacctag agatttatca gtaaacttgg tctatttcgt gtttggtatg catgttgtgc tttatccgcc agttaatagt ggccgcagtg tatgcggcga atccgtaaga cagaacttta cttaccacta atcttttact aaagggaata ttgaagcatt aaataaacaa agtgatgtat aatcatggtc tacgagccgg ccgtgctgga taattgcgtt aatgaatcgg aggcggtaat teegeeeeee caggactata cgctcactga aaggccagca cgaccctgcc ctcatagctc gtgtgcacga agtocaacco gcagagcgag acactagaag gagttggtag gcaagcagca cggggtctga caaaaaggat gtatataga cagcgatctg cgatacggga gtcctgcaac gtagttcgcc caccggctcc cacgetegte catgatcccc gatcttcagc ctgtcatgcc gagaatagtg atgccgcaaa gtatttagaa gaagtaagtt cgccacatag tctcaaggat ttcaatatta tagcaccaca cagctgcatt ttccataggc tetectgtte tcqttaaata ccacacaca taagattaat atgtgagcaa gtggcgcttt aagctgggct tatogtottg cgcttggcgt taactcacat gctcactcaa cgaaacccga aactacggct gattttggtc atgagattat aacaggatta ttcggaaaaa tttttgtt atcttttcta tcaatctaaa tagataacta gctagagtaa atcgtggtgt aggcgagtta gcacctatct gacccacgct cgcagaagtg atcgttgtca aattctctta aagtcattct gggcgaaac ggaaggcaaa ctcttccttt gataataccg gcacccaact atatttgaat gtgccac atttctgtaa atgagtgagc gctggcgttt acaaatggca gttaattgcg gctcacaatt cctgtcgtgc tgggcgctct agcggtatca aggaaagaac tcagaggtgg cctcgtgcgc ttcgggaagc cgttcgctcc atccggtaac gccagttacc aagttttaaa agccactggt gtggtggcct tagcggtggt agatcctttg ccccgtcgtg gagcggatac aatcagtgag gataccgcga aagggccgag ttgccgggaa tgctacaggc ccaacgatca cggtcctccg gtactcaacc agcactgcat gtcaatacgg acgttcttcg acccactcgt aqcaaaaca tccccgaaaa aatactcata agtcgggaaa catcatcacq ataataatca attgttatcc tttagtgagg ggggtgccta gtttgcgtat ggctgcggcg gggataacgc aggccgcgtt gacgeteaag ctggaagctc cettectece cggtgtaggt gctgcgcctt cactggcagc agttcttgaa ctctgctgaa ccaccgctgg cacgttaagg gatctcaaga attaaaaatg accaatgett ttgcctgact gtgctgcaat agccagccgg ctattaattg gctccggttc tggttatggc ttgttgccat tgactggtga tcattggaaa cttgcccggc gttcgatgta tttctgggtg cgcgcacatt ttagctcctt ggaaatgttg attgtctcat 9781 9841 0321 0381 1966 10141 .0261 9901 0021 .0081 0441 0561 10741 .0201 .0501 0621 10681 10801 19801 10981 .0921 11101 11161 11281 11341 11401 11461 11041 11221 11521 11581 11641 11701 11761 11941 11821 11881

25

30

35

15

WO 2004/067743 PCT/US2003/041269

		(PA)-Ordera-IMO) Someman (Oricon and	Y (CMI)	ACD-OLGDA				
	ORG ID NO.	Noito (pimics (the property for 1) ctdacqcqc ctqtaqcqq	tgtagcggc g	cattaagcg c	cgggtgt g		cgcagcgtga	
	61	ccqctacact	tgccagcgcc		ctcctttcgc	tttcttccct	tecttteteg	
ς.	121		cqqcatcaga		gccattgcat	acgttgtatc	catatcataa	
•	181		tatattggct			tgttgacatt	gattattgac	
	241		tagtaatcaa	ttacggggtc	attagttcat	agcccatata	tggagttccg	
	301		cttacggtaa	atggcccgcc	tggctgaccg	cccaacgacc	ccegccatt	
	361		atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	
10	421		tatttacggt	aaactgccca	cttggcagta	catcaagtgt	atcatatgcc	
	481	aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgcccagta	
	541	catgacetta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tegetattae	
	601	cateqtegate	cggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	
	661	atttccaaqt	ctccaccca	ttgacgtcaa	tgggarttg	ttttggcacc	aaaatcaacg	
15	721		aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	gtaggcgtgt	
,	781	acqqtqqqaq	gtctatataa	gcagageteg	tttagtgaac	cgtcagatcg	cctggagacg	
	841	ccatccacgo	tgttttgacc	tccatagaag	acaccgggac	cgatccagcc	teegeggeeg	
	901	_	attggaacgc	ggattccccg	tgccaagagt	gacgtaagta	ccgcctatag	
	961		cacacccctt	tggctcttat	gcatgctata	ctgtttttgg	cttggggcct	
20	1021		getteettat	gctataggtg	atggtatagc	ttagcctata	ggtgtgggtt	
ì	1081		attgaccact	ccctattgg	tgacgatact	ttccattact	aatccataac	
	1141		gccacaacta	tetetattgg	ctatatgcca	atactctgtc	cttcagagac	
	1201		tetgtattt	tacaggatgg	ggtcccattt	attatttaca	aattcacata	
	1261		ccgtcccccg	tgcccgcagt	ttttattaaa	catagcgtgg	gatetreaeg	
25	1321	-	gtacgtgttc	cggacatggg	ctcttctccg	gtagcggcgg	agcttccaca	
ì	1381		ggtcccatgc	ctccagcggc	tcatggtcgc	teggeagete	cttgctccta	
	1441		ccagacttag	gcacagcaca	atgcccacca	ccaccagtgt	gccgcacaag	
	1501	_	tagggtatgt	gtctgaaaat	gagcgtggag	attgggctcg	cacggctgac	
	1561		gacttaaggc	agcggcagaa	gaagatgcag	gcagctgagt	tgttgtattc	
30	1621		cagaggtaac	tecegttgeg	gtgctgttaa	cggtggaggg	cagtgtagtc	
	1681		tegttgetge	ეენენენენე	accagacata	atagctgaca	gactaacaga	
	1741	_	ccatgggtct	tttctgcagt	caccgtcgga	ccatgtgcga	actcgatatt	
	1801		ctctttacca	attotgcccc	gaattacact	taaaacgact	caacagctta	
	1861		gccacgcatt	acttgactgt	aaaactctca	ctcttaccga	acttggccgt	
35	1921		ccaaaqcqaq	aacaaaacat	aacatcaaac	gaatcgaccg	attgttaggt	
<b>.</b>	1981		tccacaaaga	gcgactcgct	gtataccgtt	ggcatgctag	ctttatctgt	
	2041		cgatgcccat	tgtacttgtt	gactggtctg	atattogtga	gcaaaaacga	
	2101		tacgagette	agtegeacta	cacggtcgtt	ctgttactct	ttatgagaaa	
	2161			atgttcaaag	aaagctcatg	accaatttct	agccgacctt	

ctttaaagtg ttgggacaag cttaggcatg acatgatatg ggctgcaacc ttatgattga attactggta tggaattgtg tccaatctca gccatgggtt tatctattcg cggactaggc aatcgccctg aggggatcgc tttatcatca aaagtatatt ttcctgacgg tccaagaagt ggtaaaaatc ttagctcact gcgcgcaatt ctctagaact tgacctcgag gtacatttat ttattaataq gtcaataatg gaccttatgg ggtgatgcgg ctttccaaaa gtgggaggtc tccacgctgt acggtgcatt tacataactt ggtggagtat tacgccccct tccaagtctt gtgatgctgg tcagcaactt ctaaaagcaa aacttgttaa gtcctgccta cagcaattaa aaaatcagcg cggcaaagga acttactcgt ggaaattatg tctgccttag cttatcccta gtatgttgtg taagtcgagt tcatgctgct agaaacaagg caacagttcg aaaggtgcct ccttgaatac acttaaaqct taatgatttt gtaatgtgag attacgccaa gtggcggccg tegatacege atcataatat tattgactag agttccgcgt gcccattgac gacgtcaatg atcaacggga ggcgtgtacg ggagacgcca gcggccggga atatgtcaag cccaqtacat ctattaccat cacggggatt cctatagact ctcattgtca gacttgaaaa cgttttgata tggtactggt tggaaaccta aagaggctga aaaggccgaa dcacccaaac gttcatgctc acaagggaag tacgetttgg tactcagcgt aacgtactct gttggtgacc ctgttataag taataaaac aagagaacca tatgatcccg tggcattatg aaatggttgg aaaaaattag atgaatcccc cttccggctc tatgaccatg ctccaccgcg ttgtatccat tgacattgat caagtgtatc atcaagctta ccatatatgg aacgacccc actttccatt ttagtcatcg cggtttgact tggcaccaaa atgggcggta cagatcgcct tccagcctcc gtaagtaccg caccacacca tttaggctat gaagctgggt agcggaaac atctcgctct acctaaaatc tgaaattcga aaccttccga cageteagag cagaaatcga tactcagtgc ctgatttaac ttattgataa atgaacttga aaattgatcc ggaattcgat gcttgcgggc ctacacaata cacacatggt ctgagcactg acactttatq aggaaacagc aaagctggag gtccaacatt accgccatgt ctgaccgccc caagagtgac cgggaaaagc catcttgtca attgcatacg eggggteatt agtteatage gccaataggg ggcagtacat atggcccgcc catctacqta gcgtggatag gagtttgttt attgacgcaa ccgggaccga agtgaaccgt taccgagtaa attggttgga cattgtcagc gctattggcc actcaaagac tattgtataa accacccgtc acttacctgt gccgaacgag taacatgttg ctaacacagt ggcattctgg ctcgttgacc ggtggataca cccaggcttt aatttcacac aaagggaaca aatccgttga cagacctagg agattgaaga aaaatttatt tccgggatct aaaaaacaat acaacaaaa cttgattata ccgggctgca gecegeetag ccatagtaac attgacgtca atgacggtaa cttggcagta ttccccgtgc ctgcccactt acatcaatgg acteegeeee gagetegttt atagaagaca acgtcaatgg qcqaqcattc ccatggtata tgccaaattc actcattgtc ctagcaacta aaqcqaatqc ctacgccata atgetteaac cacttccagg ctactagete tctagagcga ctttaaaaat tgcctacatc ttacggtaaa gtacaatatg tcatctagtc gaagttttgc tgaacattat gatgcctatc aggtaaacgc cattaggcac agcggataac aaccctcact aatgttaatt attaagttaa agtggatccc catcagattg attggctcat taatcaatta acggtaaatg acgtatgttc qactttccta ttttggcagt tgtcgtaaca tatataagca tttgacctcc ggaacgcgga caccccattg 2401 2461 2521 2581 2641 2701 2761 2821 2881 2941 3001 3361 3061 3121 3181 3421 3481 3541 3601 3241 3301 3721 3781 3841 3901 4081 4141 3661 3961 4021 4201 **4261** 4321 4381

25

30

35

gaccattatt gctctttgcc gtggcggtag ttcctttcca tatttcgacg tttcagaaac ttcgtgtgtg gtcaacagtt gggaggcttg gaaagccgag caagggccca ggccct; gggc caacgtgaat cggcgtggag gtgcaaggtc aacaacgccg atctcgggta gtggaggcca gatggaagac taagagtcag gcagtactcg cggccgtatt cagattcacc agttatccct aggcyccctg cttccttagc tqacaaact cttcctcttc atgcgtggtg ccgtgtggtc agggcagccc qaaccaqqtc cctctccctg ctggattcac ctgaggtcac ggggcctata gtgggttatt ccataacatg tcacatatac ctccacgcga gctcctaaca gcacaaggcc ggctgacgca tgtattctga tgtagtctga tegtgaette aaatttaata ttctggcttt aggagtcggg tggagtgggt ccgtgaaagg catttggggg cagcetecae ggggacagc cgtggaactc caggactcta ggtacgtgga aggagtacaa ccaaagccaa agctgaccaa tgctggactc cagagactga ttccacatcc taacagactg cctacatctg ccaaatcttg acagcacgta tegeegtgga ggcagcaggg cgcagaagag tgaatagcct gaccgtcagt atttacaaat agcgtgggat **дс**ддсддадс tttttggctt gcctataggt ctctgtcctt gcagctcctt ccagtgtgcc gggctcgcac gctgagttgt tggagggcag gctgacagac atcattcatc tegtteatta ttcgctttgg gtgcagctgc tgtgcagcct gggaagggc tatctgcaaa accgtctcct agcacctctg ggcacccaga gagcagtaca ctgaatggca ggctctgcac aaccactaca cattactaat tatgctgcac attatgataa aaagttgagc ctcctggggg tcccggaccc ccctgaggtc aagttcaact tcccgggatg gtgacggtgt ctacagtcct aaaaccatct cccagcgaca acgeeteeeg caccgtggac aagagcaggt tgctatactg gtatagetta cgatactttc cgtggagatt gatgcaggca ctgttaacgg tttcttcgtg gacaacagac aaccctggtc cctcatgatc tatgccaata cccatttatt tattaaacat ttctccggta tggtcgctcg agacataata cgtcggatca atttaatatt aggtacccag tagagtetee aaacacgtta taccacgggg ctcctccaag cagcagcttg ggtggacaag ccaggactgg ccccatcgag cctgccccca ctacaagacc cccaccacca ccaggctcca teeggetgte agcacctgaa aggettetat gccgcgggag ccccgaaccg acatgggctc ctgcagtcac caaggatatt ctcttatgca ctattggtga ctattggcta aggatggggt ccgcagtttt cagoggotoa cagcacaatg tgaaaatgag ggcagaagaa cgttgcggtg gcgcgccacc atatatctaa cagagccgaa gggggtccct gctgggtccg ttgatggtgg atgattcaaa tatattactg 9999cca999 ccctggcacc aggactactt tgcacacctt ccgtgccctc caccgtgccc gccacgaaga ccaagacaaa ccgtcctgca ccctcccago gcctggtcaa cggagaacaa acagcaagct tgatgcatga ataggtgatg gcaacaccaa ccaaggacac aggtgtacac accetttgg teceeegtge cgtgttccgg tgggtcttt gtgtttacct atgaatttct gacacagccg agcgtcctca tecttatget gaccactccc acaactatct gtatttttac cccatgcctc gacttaggca ggtatgtgtc ttaaggcagc aggtaactcc ttgctgccgc teggetgege gtaaagccgg gcctggatga aaaagcaaaa atctcaagag ccccgaatt teggtettee tgcctggtca aacgtggtga cacaagccca cacacatgcc ccccaaaac gtggacgtga gtgcataatg tccaacaaag cgagaaccac agcctgacct aatgggcagc ttetteetet tcatgctccg accagcggcg 4621 4681 4741 4801 4981 5041 5101 5161 5221 5281 5341 5401 5461 5521 5581 5641 5821 6421 6841 4861 4921 5701 5761 5941 6121 6181 6241 6541 6661 5881 6001 6061 6301 1989 6481 6601

15

2

20

25

30

tgacacagcc

gccgaagett

aagcgccaga

tctccgggta

ctatgaggac gacaatggcc ttactcaact cgtcctaggt gcttcaagcc gacagtggcc ctccaaacaa gtggaagtcc agtcgttaag gacagtggcc aggaaacttc gecetetgee tactgccaca ttggctactg cattatccca catatqccct cgtattacgo ttacccaact aggcccgcac aagcgttaat ccaataggcc gagtgttgtt agggrgaaaa tgtatccacc ggtcaacgag actcaatggc aaaaaaggcc ttgtctcctg aatagatagg atcaagaggg taataataca aatqattata tcaccaataa cctgtaatgc tactgagtgt tggtggtcat gagatgcatt ccagctcagg ccaagctgac ccaccacacc ccgtggagaa actactactg cctctgagga cgggagccgt cgcttgagca ttgaaggggg tgtgcagagg gettettge atttttcaag aaatggatta getteacea tatagtgagt aaccctggcg aggcaccaac tiggaaattgt aatagcgaag cattttttaa agatagggtt ccaacgtcaa tgacaagatg cagtgctcag gcttgccgat agataaataa ctcttgggtt aactaagcac gatagcagga attggtagtg ttggtaaggc gacatcactc aggcttggaa ttaaaaactt ataaatccga tcctatgagc gatgaaggtg ttcccacct caggcccctg agcggagga ctgagcctga acctgctctg ttctctggct gactcctacc ggagtggaga gaagggagca agggcccttt ggaacactga tctgaatgtg gtttcctgag tgactgggaa ggccagtgcc ttccaaagca caattcqccc cagctggcgt gaatggcgaa ggggattcat gaatagaccg cgggatcata gatgatgtgc acctaaaaga ctgtgctttc taaatcagct aacgtggact tgagcttgaa aaaaaatgcc gtgacgaat gaaggtcatc gccatcccaa accggttgca tgtttggata aattaaaaca actatctgca cctgccacaa ggccaggatc gaagtcaggc ccctgagaga ccaggtggaa ggaggtgttc ggtcactctg cagcagctac tctcataagt cgtcaaggcg gaggagga tcctcgtcac ggtcacgcat cccccgcac atctaaatgg ccttgctgca gttcaatatt tttgtattat tacaacgtcg ccctttcgc aaatttttgt taaatcaaaa ggctttttat gcccggtacc actattaaag cccactactc gggttaacat ggctattctt tgcgcagcct aaaatcatta agcttatgat catgcgaaa tctttatcqt acatcatttg tccaaatcca ggccaataac gatgactctt cagccaataa caaatqcact caggacaaac ggtaccagca cctccgggat tcagtggggc atagcagccc agtacgcggc gttatcatag ctgccccctc cactggtgtg gttcaccgcg tcctctcgtg cctctgccct ccgtcaggaa acagctgcca aaatctcaac cccagcacc gagaaatcca tcgagggggg aattcgcgtt agggcgatgg tatttaccgc taacaggtgt geegtegttt gcagcacatc tcccaacagt aaatccctta acaagagtcc gatttttacc tatcgcaata togoggaata tgagcttgag atcccaccac cccatttagt agctaaatct aaaccaataa aacagtaatg gcaccttqct tcagtgtccc tggaaggcag tatgtttatt agcaaacgac accttgacta gacagcagtg cagcccaagg aacaaggcca agcaacaaca gcgccatgac cacaaaagct cctgcagaat attgctgctt gcaagaaata agatcctgat ccagtttgga cacattcctt ctgcccatga ttgcctcccc gegeteactg taatcgcctt attttgttaa gaaatcggca ttaacttaat aatttattgc cgatcgccct accgtctatc aattaacatt tggttatgca ttttatttga tcattatatt gtaaaacgct aataacagca tccctgtaaa aggtaaagcg agatataaac cgttttttcg tttactccc 7081 7201 7261 7321 7381 7441 7501 7561 7621 7681 7741 7801 7861 7921 7981 8041 8101 8161 8221 8281 8341 8401 8461 8581 8641 8701 8521 8761 8881 8941 8821 9001 9061

2

25

30

35

15

WO 2004/067743

tcaacaaatg tattgggcgc attcatcatc acqatttctq tccgctcaca ctaatgagtg aaacctgtcg gcgagcggta gttgctggcg cccttcggga cgcaggaaag ctccctcgtg ggtcgttcgc aagtcagagg cttatccggt agcagccact gaagtggtgg gaagccagtt tggtagcggt agaagatcct agggattttg atgaagttt cttaatcagt acteceeegte aatgataccg cggaagggcc ttgttgccgg cattgctaca ttcccaacga ggcagcactg ggcgtcaata tgagtactca catgagcgga atttccccga ttgaatactc cttcggtcct aaaacgttct cctggggtgc gaaataataa tccagtcggg ccctttagtg gaaattgtta geggtttgeg caggggataa accgctgcgc ttcggctgcg aaaaggccgc atcgacgete ccctggaag ccgcctttct gttcggtgta cgccactggc cagagttctt gagatatgat aaaccaccgc aaggatctca tcagctccgg ttccgcgcac taaattaaaa tagttgcctg agtctattaa acgttgttgc cggttagctc tcatggttat ctgtgactgg tcatcattgg actcacgtta gttaccaatg ccagtgctgc accagccagc gcgtttctgg gctcttgcc ccagttcgat cacggaaatg gttattgtct ccatcatgct agcttttgtt aagtgtaaag ctgcccgctt tccacagaat caggcgtttc ggcggtgcta tttcctgtgt gcggggaga cgctcggtcg aggaaccgta ggatacctgt gttcagcccg teeggeaae ccatctggcc tcaatgcgct catcacaaa aggtatotca tttggtatct cgcagaaaaa tggaacgaaa tagatccttt tggtctgaca tcagcaataa gtcgtttggt atggcttcat gtgttatcac cacgacttat cgttcatcca gcctccatcc agtttgcgca tgcaaaaag agatgctttt cgaccgagtt caaatagggg ttaaaagtgc ctgttgagat actttcacca atttatcagg gaaaagccaa ggattggcta tataccgatc gtcatagctg cggaagcata gttgcgctca tgactcgctg aatacggtta ccctgacgag gccgcttacc cggccaacgc gcaaaaggcc ataaagatac ctcacgctgt cgaaccccc gaggtatgta aaggacagta tagctcttga ctgtctattt cccggtaaga ggagggetta tccagattta gttggccgca gcagattacg gatetteace tgagtaaact gccagttaat cccatgttg tgacgctcag aactttatcc gccatccgta gtgtatgcgg tagcagaact gatettaceg agcatcttt aaaaaaggga ttattgaagc gaaaaataaa cattaattgc caaaggcggt gacccgtaat cctcgctcac ataagtgatg cgtaatcatg acatacgage acaccgtgct attaatgaat caaaaggcca ggctccgccc ttccgacct gctgtgtgca ctacggggtc cgacaggact tttctcatag ttgagtccaa ttagcagagc gctacactag aaagagttgg tttgcaagca tatcaaaag gtggtcctgc tcagaagtaa tctcagcgat gctcaccggc tgtcacgctc aatgtattta aaagtatata ctacgatacg taagtagttc ttacatgatc ttactgtcat tctgagaata ccgcgccaca actgatcttc aaaatgccgc tttttcaata aactctcaag aaaadaccaa taatagcacc gcatcgttaa gegegettgg agctaactca tgccagctgc tetteegett tcagctcact attccacaca aacatgtgag tttttccata tggcgaaacc cgctctcctg agcgtggcgc tccaagctgg ggtaacagga accttcggaa gtgtagataa cgagacccac aactatcgtc ggttttttg gtcatgagat aaatcaatct gaggcaccta gagcgcagaa ggcatcgtgg cctaactacg ttdatcttt gaagctagag tcaaggcgag ccgatcgttg accaagtcat cgggataata tacatatttg cataattctc tcggggcgaa cgtgcaccca acaggaaggc atactcttcc 9481 9301 9361 9421 9541 9601 9661 9721 9781 9841 10201 1066 9961 10021 10081 10141 10261 10321 10501 10561 10621 10921 10381 10441 10981 10681 10741 10861 11041 11101 11281 10801 11161 11341 11221 11401 2 20 25 30 35

ttatgagaaa

atattegtga ctgttactet

gactggtctg cacggtcgtt

agtegeacta

cgatgcccat tgcgagcttc

cttatggtat

2101

tegggcaata

tgtacttgtt

ctectttege tttetteeet teettteteg gccattgcat acqttgtatc catatcataa gattattgac tggagttccg cccgcccatt attgacgtca atcatatgcc gcctggcatt atgcccagta tegetattae actcacgggg aaaatcaacg cctggagacg tccgcggccg ccgcctatag cttggggcct ggtgtgggtt aatccataac cttcagagac aattcacata gatctccacg agettecaca cttgctccta gtaggcgtgt gccgcacaag cacggctgac tgttgtattc cagtgtagtc gactaacaga actcgatatt caacagetta ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga attgttaggt tgttgacatt cccaacgacc gggactttcc agcccatata catcaagtgt tagcggtttg gtattagtca tttggcacc caaatgggcg cgtcagatcg cgatccagcc gacgtaagta ctgtttttgg ttagcctata atactctgtc attatttaca ttccattact gtagcggcgg catagcgtgg teggeagete ccaccagtgt attggggctcg cggtggaggg gcagctgagt atagctgaca ccatgtgcga taaaacgact ctcttaccga gaatcgaccg ggcatgctag attaccgcca tagtaatcaa ttacggggtc attagttcat tggctgaccg cttggcagta taaatggccc tgggcgtgga aacgccaata gtacatctac tgggaytttg tttagtgaac acaccgggac ggtcccattt ttttattaaa cccattgacg tgccaagagt gcatgctata atggtatagc tgacgatact ctatatgcca ctcttctccg tcatggtcgc atgcccacca gagcgtggag gtgctgttaa gaagatgcag accagacata caccgtcgga gtataccgtt gaattacact aaaactctca aacatcaaac ID NO:101 (pTnMCS (CMV-prepro-HCPro-Lys-CPA)) ccgctacact tgccagcgcc ctagcgcccg cttacggtaa atggcccgcc ttggctattg catgtccaac ttcccatagt tcaatgacgg aaactgccca agtacatcaa gctataggtg ctacttggca ttgacgtcaa aaatgtcgta acaactccgc tccatagaag ggattccccg cggacatggg gcacagcaca gcagageteg tggctcttat ccctattgg tctctattgg ctccagcggc gtctgaaaat tcccgttgcg tacaggatgg tgcccgcagt agcggcagaa cgcgcgcgc tttctgcagt attetgeeee acttgactgt aacaaaacat gcgactcgct cggcatcaga tatgtacatt tatattggct atgacgtatg tatttacggt cctattgacg tgggacttc cggttttggc ctccaccca attggaacgc tctgtattt ggtcccatgc gtctatataa tgttttgacc cacacccctt gcttccttat gccacaacta aagtaacaag gtacgtgttc tegttgetge attgaccact ccagacttag gacttaaggc cagaggtaac ccatgggtct gccacgcatt tagggtatgt ctctttacca ccaaagcgag tccacaaaga taqttattaa atgggtggag 11581 aaagtgccac ccacgttcgc catgacetta cgttacataa gacgtcaata ggactttcca aagtacgccc catggtgatg atttccaagt acggtgggag ccatccacgc ggaacggtgc actctatagg atacacccc attgaccatt tgacacggac tacaacaacg atggctcttt cgaatctcgg gcagatggaa tgagcagtac acagtggagg gccgtggcgg ctgttccttt acgttggctt aacctgccaa tgataagagt ttacacgact aatcotcacc tccgagccct 181 241 721 301 361 481 661 781 841 1021 1081 1141 961 1201 1261 1381 1321 1441 1501 1621 1681 1741 1801 1861 1561 1921 1981 SEO S 15 25 30 35

ttgggacaag cttaggcatg ggctgcaacc aggggatcgc ctttaaagtg tccaatctca aatcgccctg tttatcatca aaagtatatt attactggta ttagctcact tggaattgtg gagadaatt gtacatttat gtcaataatg ggtggagtat ctcgacacgg gccatgggtt tatctattcg cggactaggc ttatgattga tccaagaagt ttcctgacgg ggtaaaaatc tgacctcgag ttattaataq gaccttatgg ctctagaact tacataactt tacgccccct ggtgatgcgg tccaaqtctt acggtgcatt cagcaattaa accaatttct gtgatgctgg tcagcaactt gtcctgccta ccttgaatac taatgattt ctaaaagcaa aaaatcagcg cggcaaagga aacttgttaa ggaaattatg aaaggtgcct tctgccttag acttaaagct gtaatgtgag tcgataccgc gcccattgac taagtcgagt tcatgctgct agaaacaagg caacagttcg acttactcgt cttatcccta gtatgttgtg attacgccaa gtggcggccg atcataatat tattgactag agttccgcgt gacgtcaatg cccagtacat ctattaccat cacggggatt tggcaccaaa atcaacggga gcggccggga atatgtcaag ggcgtgtacg ggagacgcca gttcatgctc tatgaccatg ttgtatccat caagtgtatc ctcattgtca tggaaaccta aagaggctga aaaggccgaa acacccaaac gacttgaaaa cgttttgata acaagggaag tacgctftgg gttggügacc ctgttataag taataaaac aaaaaattag atgaatcccc tatgatcccg atcaagctta tgacattgat tggcattatg tccagcctcc tggtactggt tactcagcgt aacgtactct ctgatttaac aaatggttgg aagagaacca cttccggctc cagategeet ctccaccgcg ccatatatgg actttccatt cggtttgact aacgacccc ttagtcatcg agcggaaaac atgttcaaag caccacaccg gaagctgggt tttaggctat atctcgctct acctaaaatc tgaaattcga aaccttccga cagctcagag gcttgcgggc cagaaatcga ctacacaata cacacatggt cgggaaaagc tactcagtgc ttattgataa atgaacttga aaattgatcc ctgagcactg catcttgtca acactttatg ageggataae aattteaeac aggaaacage aaagctggag ggaattcgat gccaataggg atggcccgcc attgcatacg agttcatago catctacgta ccgggaccga accgccatgt ctgaccgccc ggcagtacat gagtttgttt attgacgcaa gcgtggatag agtgaaccgt tattgtataa attggttgga cattgtcagc aaagggaaca gtccaacatt tttcagagca taccgagtaa aatccgttga cagacctagg actcaaagac accacccgtc acttacctgt agattgaaga gccgaacgag taacatgttg ggcattctgg aaaatttatt teegggatet aaaaaacaat acaacaaaa cttgattata ctcgttgacc ggtggataca cccaggettt ccgggctgca gctattggcc cggggtcatt gagctcgttt ctaacacagt gcccgcctgg ccatagtaac ctgcccactt atgacggtaa cttggcagta acgtcaatgg acteegeeee atagaagaca acatcaatgg cattaggcac gtacaatatg tcatctagtc tgccaaattc cacttccagg ctttaaaaat tgcctacatc gatgcctatc agtggatccc gegageatte ccatggtata actcattgtc ctagcaacta aagcgaatgc atgetteaac gaagttttgc ctactagete tctagagcga tgaacattat aggtaaacgc aatgttaatt aaccctcact ctacgccata attaagttaa catcagattg taatcaatta acgtatgttc ttacggtaaa attgacgtca ttttggcagt caccccattq tttgacctcc attggctcat acggtaaatg gactttccta tgtcgtaaca tatataagca 2461 2341 2401 2521 2641 2701 2761 2821 2881 2941 3121 3181 3241 3301 3361 3541 2281 2581 3001 3061 3421 3481 3601 3661 3721 3781 3841 3901 3961 4021 4081 4141 4201 4261 4441 10 15 20 35 25 30

ctataggcac cacggactct atctcgggta gagccctggt gtggcggtag gatggaagac taagagtcag aacaacgccg gtggaggcca gcagtactcg tatttcgacg ttcctttcca ttogtgtgtg cagattcacc gaaagccgag gggaggettg gtcaacagtt tttcagaaac cggccgtatt agttatccct caagggccca 2666;;22266 aggcgccctg cttccttagc caacgtgaat tgacaaaact attactatta atgcgtggtg cggcgtggag ccgtgtggtc gtgcaaggtc gaaccaggtc gtgggagagc agggcagccc gaacgtcttc cctatagact gggcctata gtgggttatt ccataacatg cagagactga tcacatatac ctccacgcga ttccacatcc gctcctaaca tgtattctga tgtagtctga gcacaaggcc ggctgacgca tegtgaette ctggattcac taacagactg aaatttaata ttctggcttt aggagtcggg tggagtgggt catttggggg ccgtgaaagg cagcctccac gggcacagc cgtggaactc caggactcta ctgaggtcac tgctggactc tgaatagcct cctacatctg ccaaatcttg gaccotcagt ggtacgtgga aggagtacaa agctgaccaa tegeegtaga acagcacgta ccaaagccaa ggcagcaggg gcctataggt cattactaat ctctgtcctt atttacaaat agcgtgggat gcggcggagc gcagctcctt ccagtgtgcc gggctcgcac tttttggctt atcatccatc gtgcagctgc tgtgcagcct gctgagttgt tggagggcag gctgacagac tcgtttatta tatgctgcac tatctgcaaa attatgataa agcacctctg aaagttgagc ttcgctttgg gggaagggc accgtetect ctcctggggg gtgacggtgt ctacagtcct ggcacccaga tcccggaccc aagttcaact ctgaatggca aaaaccatct tcccgggatg cccagcgaca acagcaagct caccgtggac aagagcaggt gagcagtaca acgeeteecg tgctatactg caagagtgac gtatagetta cgatactttc tatgccaata cccatttatt ttctccggta tattaaacat tggtcgctcg tttcttcgtg cccaccacca cgtggagatt gatgcaggca ctgttaacgg agacataata cgtcggatca atttaatatt aggtacccag gacaacagac taccacgggg aaccctggtc ctcctccaag cctgaggtc tagagtetee ccaggctcca aaacacgtta ctacaadacc teeggetgte ggtggacaag cctcatgatc gccgcgggag ccccgaaccg agcacctgaa ccccatcgag aggettetat cagcagcttg ccaggactgg cctqccccca ttccccgtgc ctcttatgca ctattggcta acatgggctc ataggtgatg ctattggtga aggatggggt ccgcagttt cagoggotoa cagcacaatg ggcagaagaa cgttgcggtg ctgcagtcac atatatctaa ggggtccct tgaaaatgag gcgcgccacc caaggatatt cagagccgaa gctgggtccg ttgatggtgg atgattcaaa tatattactg aggtgtacac 9999ccaggg ccctggcacc aggactactt tgcacacctt ccgtgccctc gcaacaccaa caccgtgccc ccaaggacac gccacgaaga aceteceage gcctggtcaa ccgtcctgca ccaagacaaa cggagaacaa accetttgg tccttatgct gaccactccc gtatttttac teceeegtge gacttaggca acaactatct cgtgttccgg ccatqcctc ggtatgtgtc ttaaggcagc aggtaactcc ttgctgccgc teggetgege tgggtcttt gtgtttacct atctcaagag atgaatttct gtaaagccgg gcctggatga aaaagcaaaa gacacagccg teggtettee tgcctggtca cacaagccca ccccaaac ccccgaatt accagcggcg aacgtggtga cacacatgcc gtggacgtga gtgcataatq agcgtcctca aatgggcagc tccaacaag cgagaaccac agcctgacct ttetteetet 1681 4801 4861 4981 4921 5101 5221 5281 5701 5161 5341 5401 5461 5521 5581 5641 5761 5821 5881 6121 5941 6181 6241 6001 6061 6301 6361 6421 6481 6541 1099

2

25

30

35

WO 2004/067743

ggagcttcaa attqccaqaa cgtgacagtg accetecaaa gaagacagtg gggaggaaac tgctactgcc aagagtcgtt aacttggcta agtcgtatta gcgttaccca tgtaagcgtt gttgagtgtt atgtgtatcc cagggtcaac gttatcaaga gtgaatgatt cctctccctq catctatgag ctgttactca gaccgtccta gcagtggaag ttacatatdc ccacattatc aagaggcccg taaccastag caaagggcga gataaaaag taaaatagat cacttgtctc cgagcagtac agggccctct cttactcaat ggataataat gagatttttc agctgacaca ctgtggtggt getecagete gtgactacta ggaccaagct agaccaccac gccaggcacc cgcagaagag ctggagatgc cctcctctga tgacgcttga gcaccgtgga ttttgaagg cacggarcac tgatgtgcag gtggcttctt gcaaaatgga tregetteae ccctatagtg atatgacaag tgcttaaaaa agagettgee gaaagataaa gccctcttgg aataactaag gcaataaatc acccgggagc gaaaaccctg cgtaatagcg gaatggaaat gctcatttt ccgagatagg catcagtgct catgaaggtc atcgatagca gcattggtaa atagacatca actccaacgt caaattggta ggctctgcac aaccactaca ggccaggccc agattctctg gaagatgaag ttcagcggag ctgttcccac agtgactcct gcgggagtgg tacctgagcc catgaaggga tgggttfcct gcaggccagt tatctgtgct tgttaaatca gatgatgatg tattgagctt ttggtgacga ccagccatcc ctttgtttgg aacgctaaaa aggcaaatgc actactatct ctttcctatg atcacctgct ggaagggccc cactccgaat attttccaaa acccaattcg tcgtgactgg cgccagctgg cctgaatggc aaagaataga aagaacgtgg ctccgggatc ttaggggatt aaaacctaaa cgtaaaaaat aacaccggtt taaaattaaa gatccctgag gtgtctcata gccgaaaag aacggccagg gcagaagtca ggcccaggtg tagggaggtg ctcggtcact cccgtcaag ggccagcagc ccaggtcacg gcggaggag gtgcccccg ccttcctcgt aacatctaaa accettget ccagttcaat tgttttgtat ggggccggt ttttacaacg atcccccttt agttgcgcag gttaaattt ttataaatca tccactatta tggcccacta accaaaatca gaaagcttat atacatgcga cgcggctttt ataacatcat gaggggttaa atgggccaat gctgatgact caccagccaa tcttcttat taatccaaat gattaacagg atttcgcgga gcgatcccac attggtacca gtggttatca ccacactggt gctacagctg aatgttcacc gactcctctc cttcctctgc ataaaatctc cttccccagc tgagagaaat cttgcagcac ccttcccaac taaaattcgc gcaaaatccc ggaacaagag gcatatcgca tgctatttac gctaaaccaa gcaaacagta tgatgcatga aagcgccaga ccccaggaca gacceteegg ctatcagtgg aggetgeece cagatagcag acaagtacgc ccctcgaggg ctggccgtcg atcagggcga aatgatttt atteegteag tgaagctaaa ccctgagctt aaagcacctt aaatatgttt gacagcaaac acttaatcgc gagaattaac tctccgggta gtgtcagtgt gccaccttga tcccacaaa gccctgcag ttcgcgccat gccattgctg aaqcacattc ctgctgccca caccgatcgc tcatgctccg actgacagca ggtcagccca gccaacaagg caaagcaaca acagcaagaa cctagatcct ccattgcctc cgcgcgctca aatatttgt ggctggttat aggttttatt gggtcattat ctgtttactc acagtaaaac ttcagatata gcctggaagg gccgaaatcg gttccagttt aaaaccgtct accttaactt gccaatttat ataaataaca taatccctgt tgcaggtaaa 7141 6901 7201 7021 7081 7321 7381 7441 7501 7261 7561 7621 7681 7741 7801 7861 7921 7981 8041 8101 8161 8221 8281 8341 8401 8461 8521 8581 8701 8761 8881 8941 8641 8821 9 15 2 30 35 25

qaatactqaq atcacgattt taatcaacaa gtgagggtta ttatccgctc tgcctaatga gggaaacctg taacgcagga cgcgttgctg aaqctccctc tateactteg gtaggtcgtt cgccttatcc gcgtattggg gcggcgagcg ctcaagtcag ggcagcagcc cttgaagtgg gctgaagcca cgctggtagc tcaagaagat aaaatgaagt atgcttaatc tgcaatgata taattgttgc cggttcccaa tatggcagca tggtgagtac ttaagggatt ctgacicccc agccggaagg tgccattgct cccqqcqtca tggaaaacgt gatgtaaccc tgggtgagca atgttgaata tctcatgagc ctccttcgg caaaggettg gctattcatc gctgaaataa gttcccttta tgtgaaattg aagcctgggg ctttccagtc aatcagggga gtaaaaaggc aaaatcgacg gaggcggttt tegttegget ttcccctgg tgtccgcctt tcagttcggt ccgaccgctg tategeeact ctacagagtt aacaaaccac aaaaaggatc ccatagttgc tetgegetet aaaactcacq ttttaaatta acagttacca gcccagtgc taaaccagcc cattcagctc aagcggttag cactcatggt tttctgtgac tccaqtctat tgctcatcat gatccagttc ccagcgtttc agggttattg gcaacgttgt gttgctcttg cgacacggaa cttcctdcca ctatcaatgc atcagctttt tcactgcccg caaccatcat ctgtttcctg ataaagtgta cgcgcgggga ctgcgctcgg ttatccacag gccaggaacc gagcatcaca taccaggcgt accggatacc tgtaggtatc cccgttcagc gtaggcggtg gtatttggta tgatccggca acttggtctg tttcgttcat ttaccatctg ttgtgcaaaa agacacgact acgcgcagaa tecgeeteca ggtatggctt gtaagatgct cagtggaacg acctagatcc ttatcagcaa aatagtttgc cggcgaccga gcagtgttat actttaaaag ccgctgttga tttactttca ggaataaggg agcatttatc agtggctatt aatgaaaagc gctggattgg atgtataccg atggtcatag agccggaagc tgagttgaga ccagcaaaag actataaaga tageteaege aatcggccaa cactgactcg ggtaatacgg gatctgtcta ccccctgac cctgccgctt gcacgaaccc caacccggta agcgaggtat tagaaggaca tggtagctct gcagcagatt gtctgacgct aaggatcttc atatgagtaa acgggagga ggctccagat atcccccatg tgcaacttta ttcgccagtt ctcgtcgttt taagttggcc catgccatcc acatagcaga atattattga atagtgtatg aaggatctta ttcagcatct cgcaaaaag tegeceattt caagacccgt taaataagtg tggcgtaatc actcaaaggc accacaccgt acaacatacg tcacattaat tgcattaatg gagcaaaagg cgctttctca gtcttgagtc ggattagcag cttcctcgct acccgacagg ctgttccgac acggctacac gaaaaagagt ttgtttgcaa ataggctccg tgggctgtgt tttctacggg gattatcaaa tctaaagtat ctatctcagc taactacgat cacqctcacc gaagtggtcc tggtgtcacg gagttacatg ttgtcagaag cattctgaga ataccgcgcc gaaactctc gagtaagtag ctcttactgt ccaactgatc ggcaaaatgc tccttttca tgtaaaagac ctgtaatagc atggcatcgt attgcgcgct acaattccac gtgagctaac tcgtgccagc gtatcagete cgctcttccg aagaacatgt gcgtttttcc aggtggcgaa gtgcgctctc ggaagcgtgg cgctccaage ggtaactatc actggtaaca gttaccttcg cctttgatct tggcctaact ggtggtttt ttggtcatga tttaaatcaa agtgaggcac cgatcaaggc tettegggge gtcgtgtaga ccgcgagacc gccgagcgca cgggaagcta acaggcatcg cctccgatcg atacqqqata actegtgeac aaaacaggaa ctcatactct ctgcataatt tcaaccaagt 930I 9361 9481 9661 9781 9841 10261 9901 10001 10141 10201 10381 10441 10501 10741 1966 10021 10321 10561 10621 10681 10801 10861 10921 10981 11101 11041 11161 11281 11341 11461

23

30

35

2

11521 ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc					cac	11581 coasasotoc cac	11581
	cacatttccc	gggttccgcg	aaacaaatag	ttagaaaaat	ttgaatgtat	ggatacatat	11521

		lls81 cgaaaagtgc cac	ນອນ		:		
v	SEQ ID N	ID NO:102 (promod (CHOvep-prepro-HCPro-CPA))	:102 (pInMod (CHOvep-pi	repro-HCPro-	-CPA))	aataattaca	cacacataa
)	· <b>v</b> o		tgccagcgcc	ctagcgcccg	ctectttege		tectteteg
	121				gccattgcat	acgttgtatc	catatcataa
	181	1 tatgtacatt	tatattggct	catgtccaac	attaccgcca	tgttgacatt	gattattgac
,	241	1 tagttattaa	tagtaatcaa	ttacggggtc	attagttcat	agcccatata	tggagttccg
10	301	1 cgttacataa	cttacggtaa	atggcccgcc	tggctgaccg	cccaacgacc	cccgcccatt
	361	1 gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca
	421	_	tatttacggt	aaactgccca	cttggcagta	catcaagtgt	atcatatgcc
	481	1 aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgcccagta
,	541	1 catgacetta	tgggactttc	ctacttggca	gtacatotac	gtattagtca	tegetattae
15	109	1 catggtgatg	agsttttgga	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg
	199	1 atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg
	721	1 ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	gtaggcgtgt
	781	1 acggtgggag	gtctatataa	gcagageteg	tttagtgaac	cgtcagatcg	cctggagacg
,	841		tgttttgacc	tccatagaag	acaccgggac	cgatccagcc	tccgcggccg
20	901	1 ggaacggtgc		ggattccccg	tgccaagagt	gacgtaagta	ccgcctatag
	961	1 actctatagg	cacacccctt	tggctcttat	gcatgctata	ctgtttttgg	cttggggcct
	1021	1 atacacccc	gcttccttat	gctataggtg	atggtatagc	ttagcctata	ggtgtgggtt
	1081	1 attgaccatt	attgaccact	ccctattgg	tgacgatact	ttccattact	aatccataac
,	1141	1 atggctctt	gccacaacta	tctctattgg	ctatatgcca	atactctgtc	cttcagagac
22	1201	1 tgacacggac	tctgtatttt	tacaggatgg	ggtcccattt	attatttaca	aattcacata
	1261	1 tacaacaacg	ccgtcccccg	tgcccgcagt	ttttattaaa	catagcgtgg	gatctccacg
	1321	1 cgaatctcgg	gtacgtgttc	cggacatggg	ctcttctccg	gtagcggcgg	agcttccaca
	1381	1 tecgageeet	ggtcccatgc	ctccagcggc	tcatggtcgc	teggeagete	cttgctccta
•	1441	1 acagtggagg	ccagacttag	gcacagcaca	atgcccacca	ccaccagtgt	gccgcacaag
30	1501	l gccgtggcgg	tagggtatgt	gtctgaaaat	gagcgtggag	attgggctcg	cacggctgac
	1561	L gcagatggaa	gacttaaggc	agcggcagaa	gaagatgcag	gcagctgagt	tgttgtattc
	1621	L tgataagagt	cagaggtaac	tecegttgeg	gtgctgttaa	cggtggaggg	cagtgtagtc
	1691	L tgagcagtac	tegttgetge	ეენენენენე	accagacata	atagctgaca	gactaacaga
,	1741	ctgttccttt	ccatgggtct	tttctgcagt	caccgtcgga	ccatgtgtga	acttgatatt
35	1801	L ttacatgatt	ctctttacca	attctgcccc	gaattacact	taaaacgact	caacagotta
	1981	L acgttggctt	gccacgcatt	acttgactgt	aaaactctca	ctcttaccga	acttggccgt
	1921	l aacctgccaa	ccaaagcgag	aacaaaacat	aacatcaaac	gaatcgaccg	attgttaggt
	1981	L aatcgtcacc	tccacaaaga	gcgactcgct	gtataccgtt	ggcatgctag	ctttatctgt
	2041	L tcgggcaata	cgatgcccat	tgtacttgtt	gactggtctg	atattcgtga	gcaaaaacga

agccgacctt ctttaaagtg acatgatatg aagaggaaaa cggactaggc tccaatctca ctcgacacgg gccatgggtt tatctattcg aatcgccctg ttgggacaag cttaggcatg gttttttgtg gtgccttcct attgcatcgc ggtacctctc ggctgcaacc ataatgatco agcaaggggg tgaccaaagg ataagcagca gttggtctgc tetetetete aaaaccttat aaccaactta tcccctaatg ggctcgtatg tccaaaggag attagccttg tcccggtaat ccatgattac ccgcggtggc gcttatcgat caaaacagct acaattaatg caagttctga accaatttct gtgatgctgg taagtcgagt tcagcaactt ctaaaagcaa aaaatcagcg cggcaaagga aacttgttaa gtcctgccta tcatgctgct agaaacaagg caacagttcg ggaaattatg CCCCCCCCC aaatgaggaa ggggcagcac acttactcgt ctgtgtgttg ggctctatg cctctctctc attgacccgg agtgcctgtt gatccaagag ttaacaaatg gataataata cttgaaaaa tgtcatatga cactgatgaa ttatgcttcc acagctatga tggagctcca tegatateaa tgaactcaca cacaatccca gcaatggcaa caaacagaga aatggttgct aacaaaatct cacggtcgtt aaagctcatg aagaggctga ctcattqtca tggtactggt tggaaaccta tactcagcgt acacccaaac gacttgaaaa aaaggccgaa cgttttgata gttcatgctc aacgtactct acaagggaag tacgetttgg ctctagagat ctgttgtttg tttcctaata 9999t9999t tctctcggta gtgctgaaga aaaaactgat gggatgcggt ttggaatgaa tcagcaaatt gctttacact cacacaggaa gaacaaagc acaattactc ttatattatt tgaccctgag ctgcaggaat ggggaaacaa gcgacaaaga caaagaggag atacacatct aggtggacta accaaggtat tctctccatg atctcgctct agtegeaeta agcggaaaac atgttcaaag tttaggctat acctaaaatc tgaaattcga agattgaaga aaccttccga gattgaggga caccacaccg gaagctgggt cagctcagag ctacacaata cacacatggt aagatcagag tgccagccat tctattctgg tetetetete teggtaceag acatcacaac cagaaatcga cccactgtcc aggcatgctg attatcttga aaaataaaa ctatcattgg aacgccattg taattctcgt gttaaggtgg ataacaattt tcactaaagg tcttcaaaca aaatqatttc ggcaccccag atcccccggg gaaaaatgcc tggctacagc attgtagtgg ctcacaatat taccgagtaa tattgtataa tgcgagcttc tttcagagca cagacctagg actcaaagac aatccgttga accacccgtc acttacctgt gccgaacgag taacatgttg ggcattctgg tggctaataa gccttctagt aggtgccact agacaatagc atatttgaac gaagtgatgc taggtgtcat tetetetete catcacttta attgatgcct gacggaatgt ctaacacagt aaaatttatt totototo aaatcattaa tggtaaggta tcactcatta ttgtgagcgg gaactagtgg ccatcactga gacctgagca aatttccaca taatgtcaaa tcgagctgca tacctgattt caattaaccc cttatggtat acatteceae ccatggtata tgccaaattc ctagcaacta gcgagcattc gtacaatatg tcatctagtc actcattgtc aagcgaatgc ctacgccata atgetteaac cacttccagg gaagttttgc ctactagctc agatcacttc gatctgctgt tgaccctgga attgtctgag aggattggga tetetetete tetetetete tgccttttat attaattatg cttagaaagt ccctatccaa aatacattac attttggtaa aagctttcct gtgagttagc ttgtgtggaa gccaagcgcg ggccgctcta accgctgacc cttgacctga cagagagaa attcatctgt tgctccttcc 2161 2281 2221 2401 2581 2641 2461 2521 2701 2761 2821 2881 2941 3001 3061 3241 3301 3361 3421 3541 3121 3181 3481 3601 3661 3781 3721 3841 3901 3961 4021 4081 4141 4201 1261 1321

25

35

15

agagccttag cttcatatcc caagattttc atgaaacaaa ctttgaggaa ttgtatggag ccttccaact gcatttcata tgcactacca atatatgttt gctagtggta ctgaagaatt ataactgaaa ccagctgaat cccataaaaa ctctagtctc aaataatagg ataagcacac tctgttttct tcagaatggt tgtagataca tatttcccag cacattgtta attgaaacta ttctgaaggg agtggatcaa tttaatattt ggtacccagg agagtctcct caggctccag accacqqqqa tcctccaaga ttcttcgtgt acaacaqact cccgaaccgg aacacgttat gtctttattc aaagggcagc acaaaactat tcccaggtaa agactgacat tttccacagt ctaaacttca tatgtatgtc tttacagagg ttggtttgca agcattcaca aaaaagcagg taaaaacaaa aaatattatt gaacactcct catggaagat ctdctcttca tttgtcaaga aaccatagta cccatattt aaatggacct aaggtcaaac aatagcttct agatgcacgt acatgagcaa aggcagaaa gggctcagcc tatatctaaa aaggatattt ggactacttc agagccgaaa ggggtccctt ctgggtccgc tgatggtggg cctggcaccc tgattcaaaa atattactgt gggccaggga gacctgtgag ttgccaccta aggtctccga ctggcatagg acaatgtgtg aagcttctgc ctatcatcac aagacgtaat tccagcaaaa cagtgatctc ttaagttatt aaattgagta tttgttataa gaaagtgcgt agaatgtaga atggcactga caaacagtgc aatacajaac ctcgaaccat tttctcctat agcatggctc tgttacctac ctctaaggtt ttagcagaac tatggtgtca gctcaatgga attcaatcca tgtttaccta tgaatttctc atattcccca cggctgcgcc taaagccggg cctggatgag aaaccaaat tctcaagaga acacagccgt ccccgaattg aggtattaca gcctggtcaa gggcgatcca tataactttc tctctctaga aatgaaatgc gcattgtcaa ggaagcacat atagaagcaa ggtgtagaca acttctggtc cttagcaagg acaaaagaaa acagagattg atttttaaat ctattatttc atgattgtcc tgccaggcca aacacatttg agaaaaaag ctataggaaa tgcttcttta tgttttccat tcctgatgga tttgctctcc aaatgatttc tgtcctgcct tgtaactgaa ttcaggctat tegtgtgtgg aaagccgagg atttcgacga tcaacaqttt ggaggcttgg ttcagaaacg ggccgtatta agattcacca gttatccctc aagggcccat gccctgggct aaagtgcttt acagacccac cttgggacaa gtgctgggca acatttactg agtggctgaa tgttgcttac tttgtcaatt ctctgtcatc tgatacctgt agteeteaga aaggagagaa attctgctta ttttcccc attgtatatt tcatatcata ctgtatcctc aatattccag ttcacatgca aatgaaacag ttctcactta ttgctaatta attatctggt tgggtcacaa cgtgacttct aatttaatat tggattcact cgtgaaaggc atttggggga catccaacaq aaaactaata caatcccatt ggagtcgggg ggagtgggtc gaatagcctg gggcacagcg tctggctttg ageetecace ttttggataa tggtttaggg tcacaaaagg atgttgtact ctgacctttt ttgcacagct gcaagaagat ggtagagata agactttctt tcttgtctta taggaaagta ttctttactg tatatttgct aaaaatcagt ttcacaattc cactgcaagt ctatgttttg gatagattta tcacgtcttg agaaggatca taaaqatccc agttggctcc taactaataa tettetetee cccagaatta aaatctaacc aacctgtggg cgtttattaa tegetttggt ttatgataac tcattcatct tgcagctgca gtgcagcctc ggaaggggct atgctgcacc atctqcaaat acqtatacta gcacctctgg 4501 4561 4621 4681 4801 4861 4921 4981 5041 5101 5161 5641 5221 5281 5341 5401 5461 5521 5581 5701 5761 5821 5881 5941 6001 1909 6121 6181 6241 6301 6361 6421 6481 6541 6601 2 15 2 25 30 35

cgtgccctcc agcagcttgg cctgaggtca ccggctgtcc gtggacaaga gcacctgaac ccgcgggagg cccatcgaga ctgcccccat ggcttctatc tacaagacca accgtggaca gctctgcaca ccgaagcttt gccaggatca caggactggc aagtcaggcc cctgagagat caggtggaag ctcataagtg ctcatgatct gaggtgttca gtcactctgt gtcaaggcgg agcagctacc gtcacycatg agggaggaa cccccgcacg cctcgtcact tctaaatggg cttgctgcag ttgtattatc cccggtaccc gcgcagcctg aatttttgtt ttcaatatt acaacgtcgt cccttcgcc ctattaaaga gcacaccttt caacaccaag accgtgccca caaggacacc ccacgaagac caagacaaag cgtcctgcac cctcccagcc ggtgtacacc cctggtcaaa ggagaacaac cagcaagctc gatgcatgag agcgccagag ctccgggatc cágtggggcc ttatcatagg aggacaaacg gtaccagcag tgcccctcg actggtgtgt tagcagccc gtacgcggcc cagctgccag ttcaccgcgg cctctcgtgc aatctcaaca gatcctgatt aacaggtgtt cgagggggg ttttgttaaa attogogtta cagtttggaa caagagtcca ctctgccctt cccaqcaccc agaaatccag cagcacatcc ccgtcgtttt cccaacagtt aatcccttat tggacgtgag ccagcggcgt acgtggtgac cccaaaacc tgcataatgc gcctgacctg acaagcccag acacatgccc gcgtcctcac ccaacaagc gagaaccaca atgggcagcc tcttcctcta catgetregt ctccgygtaa cagtgtcccc acagcagtgg gtcctaggtc agcccaaggc cttcaagcca acaaggccac atgtttattg gcaaacgacc ccttgactat ggaaggcaga acaaaagcta ctgcagaatg ttgctgcttc caagaaataa tgcctcccct gcaacaacaa cgccatgact acattccttc tgcccatgag cgctcactgg aaatcggcaa aatcgccttg gategeeett ggcgccctga ttccttagca aacgtgaatc gacaaaactc tgcgtggtgg ggcgtggagg ttcctcttcc cgtgtggtca tgcaaggtct tgggagagca gggcagccc aaccaggtca ccagaaaat gacggctcct aacgtcttct ctctccctgt ccctcggtgt tatgaggaca acaatggcca tactcaactg acagtggcct acagtggccc tccaaacaaa tggaagtccc actgccacag gtcgttaagc ggaaacttcg ccctctgcca tggctactgc atatgcccta attatcccat gtattacgcg tacccaactt ggcccgcacc agcgttaata caataggccg agtgttgttc ctacatctgc gtggaactca aggactctac caaatcttgt accgtcagtc tgaggtcaca gtacgtggac cagcacgtac cgccgtggag ggagtacaag caaagccaaa gctgaccaag gctggactcc cageteaggg gcagcagggg gcagaagc ggtggtcatc gacacagcca agatgcattg ctactactgt caagctgacc ctctgaggag gggagccgtg caccacaccc gcttgagcag cgtggagaag tgaaggggga gtgcagaggg cttctttgct ttttcaaga aatggattac cttcacccac aaatcagctc attttttaac atagtgagtc ggcaccaact atagcgaaga aatagaccga gatagggttg accctggcgt ggaaattgta tacagteete aagttgagcc tcctgggggg gcacccagac cccggacccc agttcaactg agcagtacaa aaaccatctc cccgggatga cctgctctgg tgaatggcaa accactacac tctctggctc atgaaggtga gcggaggac ccagcgacat agactacat agagcaggtg cctatgagct tcccacctc aggcccctgt actectacee gagtggagac tgagcctgac aagggagcac gggcccttt gaacactgat ctgaatgtgg tttcctgaga gactgggaaa agctggcgta gccagtgcca tccaaagcaa tgtgctttca aatggcgaat aattcgcct 7141 7081 7261 7321 7381 7201 7441 7501 7621 7741 7561 7681 7801 7981 8041 8101 7861 7921 8161 8221 8281 8521 8341 8401 8461 8581 8641 8701 8761 8821 8881

25

30

aaatcattag ggttaacatg gattggctat ttgcgctcac ccactactcc gcttatgatg atgcgaaaaa gctttttatt ctttatcgta catcatttgg ccaaatccad gccaataaca agccaataaa aaatgcacta gctattcttc aaaagccaac ataccgatca ggaagcataa gactogetge cctgacgagc atgactctt tcatagctgt atacggttat taaagatacc ggccaacgcg caaaaggcca ccgcttaccg tcacgctgta gaacccccg aggtatgtaq agctcttgat cagattacgc gagtaaactt ccggtaagac aggacagtat gacgeteagt atcttcacct gaggcttac gggcgatggc attttacca atttaccgcg ctaaaaaggc cgcggaataa gagcttgagg teceaceace ccatttagtg cgtcaggaaa atcgcaatac gctaaatctt aaccaataat acagtaatgg caccttgctg accegtaatg caccgtgctg taagtgatgt gtaatcatgg catacgagco attaattgcg ttaatgaatc cageteacte aaaggeggta acatgtgagc aaaaggccag geteegeeee gacaggacta tgagtccaac aagagttggt ctcgctcact teegaceetg ttctcatagc ctgtgtgcac tagcagagcg ctacactaga tacgatacgg ttgcaagcag atcaaaaagg tacggggtct aagtatatat taacttaatg ggttatgcat atttattgct tttatttgaa ataacagcaa ggtaaagcga ccgtctatca attaacattc cattatattt ttactccct taaaacgcta ccctgtaaag gatataaacg gttttt⁺cgc gctaactcac gccagctgca aaagaccaag catcgttaaa ttccacacaa cttccgcttc aatagcacca agadattgga ttttccatag ggcgaaaccc gctctcctgt gcgtggcgct ccaagctggg gtaacaggat ctaactacgg ccttcggaaa gttttttgt tgatcttttc tcatgagatt aatcaatcta aggcacctat tgtagataac actatogtot gggcgaaaa gtatccacct gtcaacgaga ctcaatggct aaaaaggcca atagataggt tcaagagggt tgtctcctgt aataatacag atgattataa ctgtaatgca ccaaccttca cagtactgcc actgagtgta cgatttctgt caacaaatgg caccaataat ccgctcacaa taatgagtga gggttaattg attgggcgct cgagcggtat gcaggaaaga ttgctggcgt tecetegtge aacctgtcgt ccttcgggaa ttatccggta gcagccactg aagccagtta tgaagtttta agtcagaggt gtcgttcgct aagtggtggc ggtagcggtg gaagatcctt gggattttgg ttaatcagtg ctccccgtcg taaaaactta cttgccgata gataaataaa tcttgggtta ttggtagtga ggaaaactaa caacqtcaaa gacaagatgt agtgctcagg actaagcact atagcaggat tggtaagget acatcactcc taaatccgag ggcttggaat ttcatcatca cctttagtga aaattgttat ctggggtgcc aaataat ccagtcggga cggtttgcgt teggetgegg aggggataac tegaegetea ccctggaagc agactttata gccactggca agagttcttg cgctctgctg aaattaaaa ttaccaatge aaaggccgcg ttcggtgtag ccgctgcgcc aaccaccgct aggatctcaa ctcacqttaa agttgcctga acgtggactc gggattcatc gggatcatat atgatgtgct aaaaatdccc tgacgaaata ccatcccaaa gtttggatag attaaaacag ttcctgtgtg cctaaaagag gagcttgaaa aaggtcatcg ccggttgcat ctatctgcaa ctgccacaa caatgcgctg catcatgcta gcttttgttc agtgtaaagc tgcccgcttt ccacagaatc ggaaccgtaa atcacaaaaa aggcgtttcc ttggtatctg agatcctttt cggggagagg gctcggtcgt gatacctgtc acqacttatc geggtgetac gcagaaaaa ggaacgaaaa ggtctgacag ggtatctcag ttcagcccga ccggcaaaca gttcatccat 9181 9241 9301 9361 9421 9481 9541 1096 9841 9901 10201 9661 9721 9781 1966 10021 10081 10141 10261 10321 10381 10441 10501 10561 10621 10681 10741 10801 10861 10981 10921 11101 11041 11221 11281 11401 11161

2

25

30

15

tat ccg gta tgt cag ctt tta	gca sac	.ga kaa jac cg ca	្រី	משש ת משמ משמ ת ני משמ ה
ccagatttat actttatccg ccagttaata tcgtttggta cccatgttgt ttggccgcag ccatccgtaa tgtatgcggc agcagaactt atcttaccgc gcatcttta	tattgaagca aaaaataaac	cgcagcgtga tcctttctcg catatcataa gattattgac tggagttccg cccgcccatt	atcatatgoc atgoccagta togotattac actcacgggg aaaatcaacg gtagggggtgt	cogectatag cttggggcct ggtgtgggcct aatccataac cttcagagac aattcacata
ctcaccggct tggtcctgca aagtagttcg gtcacgctcg tacatgatcc cagaagtaag tactgtcatg ctgagaatag ctgagaatag cgcgccacat actctcaagg ctgagaatag	ttttcaatat atgtatttag	ggtggttacg tttcttccct acgttgtatc tgttgacatt agcccatata cccaacgacc	gcctggcatt gtattagtca tagcggtttg tttttggcacc caaatgggcg cgtcagatcg	
	ractcttcct acatatttga aagtgccac	-LVS-CPA)) ctctttcgc gccattgcat attaccgcca attactgcct attactgcct attagttcat tggctgaccg	taaatggccc gtacatctac gggcgtgga tgggagtttg cccattgacg	
atgataccgc ggaagggccg tgttgccggg attgctacag tcccaacgat ttcggtcctc gcagcactgc gagtactcaa aaacgttctt taacccactc	tgaatactca atgagcggat tttccccgaa	epro-HCPro- gcattaageg ctagegeceg ttggetattg catgtecaae ttaeggggte ttgecegec atggecegec	tcattgacgg ctacttggca agtacatcaa ttgacgtcaa acaactccgc gcagagctcg	
cagtgctgca ccagccagcc gtctattaat cgttgttgcc cagctccggt ggttagctcc catggttatg tgtgactggt ccttgcccg catcattgga catcattgga cattctgaga	acyyaaatyt ttattytoto toogogoaca	(CHOvep-pr ctgtagcggc tgccagcgcc cggcatcaga tatattggct tatattggct tatataggct actacggtaa	cotattgacg tgggactttc cggttttggc ctccaccca aaatgtcgta gtctatataa	
	tattattaggggt aaataggggt	NO:103 (pTnMod (CHOvep-prepro-HCPro-LYS-CPA))  1 etgacgcgcc ctgtagcggc gcattaagcg cggcggtg 61 ccgctacact tgccagcgcc ctagcgcccg ctccttccg 21 ccacgttcgc cggcatcaga ttggctattg gccattgca 81 tatgtacatt tatattggct catgtccaac attaccgcc 841 tatgtacatta tatattggct attacggggtc attagtccaac 841 tagttactaa tagtaatcaa ttacggggtc attagtcca 842 tagttacataa cttacggtaa atggcccgcc tggctgacc 843 gacgtcaata atgacgtaa ttcccatagt aacgccaat 844 aacgccaat 845 aacgccaat		<del>-</del>
11461 11521 11581 11641 11701 11701 11881 11881 11941 12001 12121	12241 12341 12301	SEQ ID NO: 1121 121 121 241 361		901 961 1021 1081 1141 1201 1261
5 10	15	20	30	35

WO 2004/067743 PCT/US2003/041269

cttgctccta tgttgtattc cagtgtagtc gactaacaga ctttaaagtg caacagctta gcaaaaacga ttatgagaaa aagaggaaaa acatgatatg tccaatctca ctcgacacgg cggactaggc cttaggcatg acttgatatt acttggccgt attgttaggt ctttatctgt agccgacctt gccatgggtt tatctattcg gtttttgtg aatcgccctg ttgggacaag ggctgcaacc ataatgatcc gtgccttcct tgaccaaagg ataagcagca attagccttg attgcatcgc agcaaggggg ggtacctctc tctctctctc gttggtctgc gaatcgaccg ggcatgctag atattogtga teggeagete ccaccagtgt attgggctcg gcagctgagt cggtggaggg atagctgaca ccatgtgtga taaaacgact ctcttaccga ctgttactct accaatttct gtgatgctgg taagtcgagt tcagcaactt ctaaaagcaa aaaatcagcg cggcaaagga aacttgttaa gtcctgccta tcatgctgct ggaaattatg agaaacaagg aaatgaggaa ggggcagcac gggctctatg actatatata caacagttcg ctgtgtgttg ccctcccc ttaacaaatg gataataata cttgaaaaa acttactcgt attgacccgg agtgcctgtt ctetteteeg gagcgtggag gaagatgcag gtgctgttaa accagacata aagaggctga tcatggtcgc atgcccacca caccgtcgga gaattacact aaactctca aacatcaaac gtataccgtt gactggtctg aaagcccatg ctcattgtca aaaggccgaa acacccaaac cacggtrgtt tggtactggt cgttttgata tggaaaccta tactcagcgt gttcatgctc aacgtactct tctctcggta gacttgaaaa acaagggaag tacgetttgg ctctagagat ctgttgtttg tttcctaata gggatgcggt gtgctgaaga aaaaactgat ttggaatgaa 9999t9999t acaattactc ttatattatt cggacatggg ctccagcggc gcacagcaca gtctgaaaat agcggcagaa tcccgttgcg သစ်သစ်သစ်သစ်သ tttctgcagt attetgeece acttgactgt aacaaacat gcgactcgct tgtacttgtt agtogoacta atgttcaaag agcggaaaac atctcgctct caccacaccg gaagctgggt tttaggctat acctaaaatc tgaaattcga cagctcagag aaccttccga gcttgcgggc aagatcagag cagaaatcga ctacacaata tgccagccat ctatcattgg cacacatggt ccactatcc tctattctgg aggcatgctg tetetetete teggtaceag aaaataaaa acatcacaac attatcttga ccagacttag cagaggtaac tegttgetge tccacaaaga ggtcccatgc tagggtatgt gacttaaggc ccatgggtct ctctttacca ccaaagcgag tttcagagca taccgagtaa gccacgcatt cgatgcccat tgcgagcttc aatccgttga actcaaagac tattgtataa accacccgtc agattgaaga cagacctagg gccgaacgag taacatgttg gccttctagt aggtgccact acttacctgt tggctaataa taggtgtcat agacaatagc tetetetete atatttgaac gaagtgatgc ctaacacagt ggcattctgg aaaatttatt tetetetete catcacttta attgatgcct gcagatggaa ttacatgatt acagtggagg gataagagt tgagcagtac ctgttccttt aacctdccaa aatcgtcacc tegggeaata gegtteeege gcgagcattc tccgagccct gccgtggcgg acgttggctt cttatggtat ccatggtata gtacaatatg tcatctagtc tgccaaattc actcattgtc ctagcaacta aagcgaatgc ctacqccata atgetteaac cacttccagg gaagttttgc attgtctgag aggattggga ctactagete agatcacttc gatctgctgt tgaccctgga tetetetete tetetetete tgccttttat attaattatg cttagaagt ccctatccaa 1381 1501 1681 1861 1621 1741 1801 1921 2101 1981 2041 2161 2281 2461 2521 2581 2701 2221 2341 2401 2641 2761 2821 2881 2941 3001 3061 3121 3181 3241 3301 3361 3421 3481

2

25

30

35

2

tccaaaggag ggctcgtatg ccatgattac gcttatcgat caaaacagct tcccggtaat tcttccctc acaattaatq caagttctga cttcatatcc agagccttag ttgtactgct gcatttcata atgagactgt ccttccaact atgaaacaaa tgcactacca ataactgaaa caagattttc atatatgttt gctagtygta ctgaagaatt tcagaatggt ccagctgaat ctttgaggaa ttgtatggag cccataaaaa ataagcacac aaataatagg tgtagataca attgaaacta cactgatgaa tgtcatatga ttatgcttcc acagetatga tggagctcca tegatateaa tgaactcaca cacaatccca gcaatggcaa aatggttgct ctcacctgga caaacagaga gtctttattc tcccaggtaa aacaaaatct aaagggcagc acaaactat agactgacat tttccacagt aaaaagcagg taaaaacaaa aaatattatt ctaaacttca tatgtatgtc tttacagagg ctattatttc aatacagaac aatagcttct ctcgaaccat gaacactcct aadacatttg aaattgagta ttggtttgca tttgttataa agcattcaca ttaagttatt catggaagat tttctcctat tttgtcaaga tgttttccat ctctaaggtt cccatatttt gctcaatgga acatgagcaa ttagcagaac aggcagaaaa gaaagtgcgt ctgctcttca tgtcctgcct agcatggctc agatgcacgt tgttacctac aaccatagta tttgctctcc attcaatcca aaatggacct cacacaggaa gaacaaagc tcagcaaatt tctctccatg tgaccctgag atacacatct gctttacact ctgcaggaat aggtggacta ggggaaacaa accaaggtat gcgacaaaga caaagaggag ttgccaccta gacctgigag aggteteega ctggcatagg acaatgtgtg aagcttctgc ggaagcacat ctatcatcac tccagcaaaa cttagcaagg agaatgtaga cagtgatctc caaacagtgc aagacgtaat atggcactga ccatcactga tggctacagc cttgggacaa gcattgtcaa aacgccattg taattctcgt gttaaggtgg ggcaccccag tcactaaagg gaaaaatgcc aaatgatttc attgtagtgg acagacccac aatgaaatgc ataacaattt atcccccggg tcttcaaaca tataactttc gggcgatcca ctcacaatat aagaggcttt tctctctaga atagaagcaa ggtgtagaca acaaaagaaa attotgotta acagagattg atgattgtcc tgccaggcca acttctggtc ctataggaaa tgcttcttta tgtaactgaa atttttaaat agaaaaaag tcctgatgga tggtaaggta gaactagtgg tacctgattt gacctgagca taatgtcaaa gtgattggat aaagtgcttt gtgctgggca gacggaatgt aaatcattaa tcactcatta ttgtgagcgg caattaaccc tegagetgea aatttccaca tgttgcttac acatttactg agtggctgaa tgatacctgt agtcctcaga aaggagagaa ttttcccc tttgtcaatt attgtatatt tcatatcata ctgtatcctc aatattccag ctctgtcatc ttctcactta ttcacatgca aatgaaacag ttgctaatta attatctggt catccaacag aaaactaata attttggtaa ttgtgtggaa aagctttcct gtgagttagc gccaagcgcg ggccgctcta accgctgacc cttgacctga attcatctgt ttttggataa cagagagaa atgaaaaggc tgeteettee gtaggtttta tggtttaggg ttgcacagct ggtagagata ctgacctttt gcaagaagat taggaaagta agactttctt aaaaatcaqt tcttgtctta tcacaaaagg atgttgtact ttctttactg ttcacaattc cactgcaagt gatagattta agttggctcc tatatttgct ctatgttttg tcacgtcttg taactaataa agaaggatca taaagatccc tettetetee cccagaatta 3841 390I 3961 4141 4021 4081 4261 4321 4381 444I 4201 4501 4561 4741 4621 4681 4801 **4861** 4921 4981 5161 5221 5281 5341 5041 5401 5461 5521 5581 5641 5101 5701 5761

135

25

30

35

15

ttcttcgtgt tttaatatt ggtacccagg agagtctcct caggctccag accetggtea cccgaaccgg gtggacaaga cctgaggtca acaacaqact aacacgttat tcctccaaga agcagcttgg gcacctgaac ctcatgatct ccgcgggagg caggactggc ctgccccat accacggggs ccggctgtcc cccatcgaga ggcttctatc tacaagacca accgtggaca gctctgcaca ccgaaaaagc acggccagga cagaagtcag atccctgaga gcccaggtgg agggaggtgt togatoacto tgtctcataa cccgtcaagg cggagggagg aaqqtcaaac gggctcagcc tatatctaaa aaggatattt ggggtccctt tgatggtggg atattactgt cgtgccctcc agagccgaaa gggccaggga cctggcaccc ggactacttc gcacaccttt ccacgaagac ctgggtccgc tgattcaaaa caacaccaag caagacaaag accgtgccca caaggacacc cgtcctgcac cctcccagcc ggtgtacacc cctggtcaaa ggagaacaac cagcaagctc gatgcatgag cccaggacaa ttggtaccag tatcagtggg agcgccagag accctccggg tggttatcat ggctgccccc agatagcagc ctacagetge cacactggtg caagtacgcg atgttcaccg aggatgaga tatggtgtca tgtttaccta tgaatttctc taaagccggg cctggatgag tctcaagaga ccccgaattg atattcccca acacagccgt aaagcaaaat cggtcttccc gcctggtcaa ccagcgjcgt acgtggtgac aacgtgaatc acaagcccag tggacgtgag ccaacaaagc gcctgacctg acacatgccc cccaaaacc tgcataatgc gegteeteac gagaaccaca tcttcctcta atgggcagcc catgctccgt ctccgggtaa tgtcagtgtc acagcaaacg aatatgttta ccaccttgac ctgacagcag gtcagcccaa ccaacaaggc cctggaaggc aaagcaacaa cccacaaag ccctgcaga aaatgattto ttcaggctat tegtgtgtgg atttcgacga ttcagaaacg ggccgtatta agattcacca gttatccctc tcaacagttt ggaggettgg aaagccgagg aagggcccat ggcgccctga ttccttagca gccctgggct gacaaaactc ttcctcttcc tgcgtggtgg ggcgtggagg cgtgtggtca tgcaaggtct gggcagcccc aaccaggtca tgggagagca gacggctcct aacgtettet ctctccctgt ccaccctcgg ttgccagaaa atctatgagg gggacaatgg tgttactcaa accgtcctag gagetteaag ccctccaaac cagtggaagt aagacagtgg gtgacagtgg caatcccatt tgggtcacaa cgtgacttct aatttaatat tetggetttg ggagtcgggg cgtgaaaggc gaatagcctg atttggggga agcetecace ctacatctgc tggattcact ggagtgggtc aggactctac gggcacagcg accgtcagtc tgaggtcaca gtggaactca caaatcttgt gtacgtggac cagcacgtac ggagtacaag caaagccaaa gctgaccaag cgccgtggag gctggactcc gcagcagggg gcagaagc tgtggtggtc gctgacacag tggagatgca tgactactac ctccagctca gaccaagctg ctcctctgag gacgcttgag cccgggagcc gaccaccaca caccgtggag aaatctaacc tegetttggt aacctgtggg tcattcatct cgtttattaa tgcagctgca gtgcagcctc ggaagggct ttatgataac agtateata atgctgcacc atctgcaaat gcacctctgg tgacggtgtc tacagtcctc gcacccagac aagttgagcc tectgggggg cccggacccc agttcaactg cccgggatga agcagtacaa tgaatggcaa aaaccatctc accactacac tttcctatga tcacctgctc ccagcgacat cgcctcccgt agagcaggtg gccaggcccc gattetetgg aagatgaagg tcagcggagg tgttcccacc gtgactccta cgggagtgga atgaaggaa acctgagcct 6121 6181 6241 6301 6361 6481 6541 6421 6601 6841 6661 6721 6781 6901 1969 7021 7081 7141 7201 7261 7561 7321 7381 7441 7681 7501 7621 7741 7801 7981 8041 7861 7921 8101 8161

2

25

30

35

cttcctcgtc cccttgctg cagttcaata gttttgtatt gggcccggta tttacaacgt tacacattta gttgcgcagc ttaaatttt tataaatcaa ccactattaa ggcccactac ccaaaatcat aaagcttatg tacatgcgaa geggetttt aggggttaac cttctttatc taacatcatt ctgatgactc aatccaaatc tgggccaata accagucaat ctggattggc ggcaaatgca gtggctattc atgaaaagcc tgtataccga gccggaagca tggtcatago atcggccaac gcgttgcgct actgactcgc ctataaagat ttectetgee ttccccagca actecteteg taaaatctca gagagaaatc attaacaggt cctcgagggg tggccgtcgt ttgcagcaca cttcccaaca aaaattcgcg atgattttta treegteagg caaaatcct gaacaagagt tcagggcgat catatogcaa gctatttacc tttcgcggaa gaagctaaat cctgagcttg caaacagtaa ctaaaccaat aagcaccttg cgateceaee acgctaaaaa cgcccattta aagacccgta ccacaccgtg aaataagtga caacatacga ctcaaaggcg ggcgtaatca cacattaatt tteetegete agcaaaaggc taggctccgc cccgacagga gcattaatga tcgcgccatg agcacattcc ccattgctgc cagcaagaaa tgctgcccat ctagatcctg cattgcctcc cttaatcgcc atattttgtt gegegeteae accgatcgcc ttccagtttg ccttaactta ccgaaatcgg aaaccg.cta agaattaaca gctggttatg ccaatttatt ggttttattt ggtcattata tgtttactcc cagtaaaacg taaataacag aatccctgta gcaggtaaag gccgttttt gtaaaagacc tgtaatagca tggcatcgtt tcagatataa caattccaca ttgcgcgctt tgagctaact cgtgccagct getetteege aacgcaggaa agaacatgtg tatcagctca cgtttttcca ggtggcgaaa ggaggaaact gggccctctg agagtcgtta tacatatgcc gtcgtattac agaggcccgc gctactgcca acttggctac cacattatcc gtaagcgtta ttgagtgttg tatgacaaga tgtgtatcca cgttacccaa aaccaatagg aaagggcgaa agggtcaacg ttatcaagag ttactcaatg ataaaaagg aaaatagata gctcaccaat acttgtctcc gataataata tgaatgatta tccctgtaat taaccaacct tcacgatttc gagcagtact tgagggttaa tatccgctca aatactgagt aatcaacaaa ggaaacctgt cgtattgggc gcctaatgag cggcgagcgg qegttgetgg tcaagtcaga ttttgaaggg caaaatggat tegetteace gatgtgcaga tggcttctt agatttttca ccaggcacca cctatagtga aaaaccctgg gtaatagcga cgagataggg ctccaacgtc atcagtgete aatggaaatt ctcatttttt gcttaaaaac aaagataaat gagettgeeg ccctcttggg ataactaagc cagggaaaac tcgatagcag cattggtaag aaattggtag tagacatcac caataaatcc ttccctttag tttccagtcg aaaggcttgg ctattcatca ctgaaataat gtgaaattgt agcctggggt aaatcgacgc aggcggtttg cgttcggctg atcaggggat taaaaaggcc gaagggccct actctgaatg acqqaacact cccaattcgc cgtgactggg gggtttcctg caggccagtg ttttccaaag atctgtgctt gccagctggc gttaaatcag tccgggatca taggggattc atgatgatgt attgagcttg ctgaatggcg aagaatagac agaacgtgga gtaaaaatg tggtgacgaa cadccatccc aaacctaaaa atgaaggtca tttgtttgga acaccggttg aaaattaaaa ctactatctg ttcctgccac tatcaatgcg tcagcttttg cactgcccgc aaccatcatg tgtttcctgt taaagtgtaa 9cgcgggag tatccacaga ccaggaaccg agcatcacaa tgcgctcggt 8401 8461 8521 8641 8581 8701 8761 8821 8941 1906 9121 8881 900I 9241 9301 9481 9181 9361 9421 9541 1096 9661 9721 9781 9841 9901 10081 9961 10141 10201 10261 10021 10381 10321 10441 10501

20

25

30

35

	10681	1 accaggcgtt		ageteeteg	tgagatataa	tgttccgacc	ctgccgctta	
	10741	1 ccggatacct	gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	
	10801	1 gtaggtatct	cagttcggtg	taggtcgttc	gctccaagct	gggctgtgtg	cacgaacccc	
L	10861	1 ccgttcagcc		gccttatccg	gtaactatcg	tcttgagtcc	aacccggtaa	
n	10921	1 gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	
	10981	1 taggcggtgc	tacagagttc	ttgaagtggt	ggcctaacta	cggctacact	agaaggacag	
	11041	1 tatttggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	
	11101	1 gatccggcaa	acaaaccacc	gctggtagcg	gtggttttt	tgtttgcaag	cagcagatta	
-	11161		aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	
2	11221	l agtggaacga	aaactcacgt	taagggattt	tggtcatgag	attatcaaaa	aggatcttca	
	11281		tttaaattaa	aaatgaagtt	ttaaatcaat	ctaaagtata	tatgagtaaa	
	11341	1 cttggtctga	cagttaccaa	tgcttaatca	gtgaggcacc	tatctcagcg	atctgtctat	
	11401	_	catagttgcc	tgactccccg	tcgtgtagat	aactacgata	cgggagggct	
15	11461		ccccagtgct	gcaatgatac	cgcgagaccc	acgeteaceg	gctccagatt	
2	11521		aaaccagcca	gccggaaggg	ccgagagaaa	aagtggtcct	gcaactttat	
	11581		ccagtctatt	aattgttgcc	gggaagctag	agtaagtagt	tegecagtta	
	11641	l atagtttgcg	caacgttgtt	gccattgcta	caggcatcgt	ggtgtcacgc	tegtegtttg	
	11701	-	attcagctcc	ggttcccaac	gatcaaggcg	agttacatga	tececatgt	
ć	11761		ageggttage	tacttaggta	ctccgatcgt	tgtcagaagt	aagttggccg	
97	11821		actcatggtt	atggcagcac	tgcataattc	tcttactgtc	atgccatccg	
	11881	. taagatgett	ttctgtgact	ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	
	11941	_	ttgctcttgc	ccggcgtcaa	tacgggataa	_	catagcagaa	
	12001	ctttaaaagt	gctcatcatt	ggaaaacgtt	cttcggggcg a	aaactctca	aggatettae	
40	12061	cgctgttgag	atccagttcg	atgtaaccca		caactgatct	tcagcatctt	
Ç	12121	. ttactttcac	cagcgtttct	gggtgagcaa	aaacaggaag i	gcaaaatgcc 🤅	gcaaaaagg	
	12181	_	gacacggaaa	tgttgaatac			tattattgaa	
	12241	gcatttatca	gggttattgt	ctcatgagcg	gatacatatt 1	tgaatgtatt 1	tagaaaaata	
	12301	aacaaatagg	ggttccgcgc	acatttcccc	gaaaagtgcc a	ac		
30	SEQ ID NO	ID NO:104 (pTnMOD (CMV-prepro-HCPro-CPA))	(CMV-prepr	o-HCPro-CPA	<b>~</b>			
	ч	ctgacgcgcc	ctgtagcggc	gcattaagcg	gegggtgt	ggtggttacg o	cgcagcgtga	
	61	ccgctacact		ctagegeeeg	ctecttege t	tttcttccct t	tectteteg	
	121	ccacgttcgc				acgttgtatc c	catatcataa	
35	181	tatgtacatt					gattattgac	
Ç	241	tagttattaa					tggagttccg	
	301	cgttacataa	-				cccccatt	
	19F	gacgreaara		-			attgacgtca	
	424	acgggragag	-	_		-	atcatatgec	
	3 H	aaycacyccc	ccrarridacy	rcaargacgg 1	raaaryycco g	gcctggcatt a	argeceagra	

aaaatcaacg tccgcggccg ccgcctatag cttggggcct ggtgtgggtt aatccataac cttcagagac gatctccacg cctggagacg aattcacata agettecaca cttgctccta tgttgtattc gccgcacaag acttgatatt caacagetta acttggccgt attgttaggt cacggctgac cagtgtagtc gactaacaga gcaaaaacga ttatgagaaa agccgacctt ctttaaagtg acatgatato aagaggaaa tccaatctca ctcgacacgg cttaggcatg ctttalctgt gccatgggtt tatctattcg cggactaggc gtattagtca tagcggtttg ttttggcacc caaatgggcg cgtcagatcg cgatccagcc ctgtttttgg gacgtaagta atactetgte ttagcctata ttccattact attatttaca catagcgtgg gtagcggcgg teggeagete ccaccagtgt attgggctcg gcagctgagt cggtggaggg ccatgtgtga ctcttaccga gaatcgaccg tagaacgact gactggtctg atattcgtga ctgttactct atgttcaaag aaagctcatg accaatttct gtgatgctgg tcagcaactt aaaatcagcg atagctgaca ggcatgctag ctaaaagcaa taagtcgagt cggcaaagga aacttgttaa gttcatgctc agaaacaagg gtcctgccta tcatgctgct caacagttcg tttagtgaac acaccgggac gtacatctac tgggcgtgga tgggagtttg cccattgacg tgccaagagt gcatgctata atggtatagc ggtcccattt ttttattaaa tgacgatact ctatatgcca ctcttc+ccg atgeceaeca tcatgytcgc gagcgtggag gaagatgcag gtgctgttaa caccgtcgga aacatcaaac accagacata gaattacact aaaactctca cacggtcgtt gtataccgtt ctcattgtca tggtactggt tggaaaccta tttaggctat aagaggctga aaaggccgaa tactcagcgt acacccaaac gacttgaaaa cgttttgata aacgtactct gcagagctcg ggattccccg ctacttggca cggttttggc agtacatcaa acaactccgc tccatagaag gctataggtg ttgacgtcaa tggctcttat ccctattgg tctctattgg tacaggatgg tgcccgcagt cggacatggg ctccagcggc gcacagcaca agcggcagaa აანანანანა attetgeece acttgactgt gtctgaaat tecegttgeg tttctgcagt gcgactcgct tgtacttgtt agtcgcacta aacaaacat caccacaccg gaagctgggt agcggaaaac atctcgctct tgaaattcga acctaaaatc aaccttccga cagaaatcga cageteagag tgggactttc aaatgtcgta ctccaccca gtctatataa tgttttgacc cacacccctt gcttccttat gccacaacta tctgtattt attggaacgc ggtcccatgc attgaccact ccgtcccccg gtacgtgttc gacttaaggc ccagacttag tagggtatgt cagaggtaac tegttgetge ccatgggtct ctctttacca gccacgcatt ccaaagcgag tccacaaaga tgcgagcttc tttcagagca taccgagtaa cgatgcccat actcaaagac aatccgttga agattgaaga cagacctagg tattgtataa accacccgtc acttacctgt gccgaacgag ctaacacagt ggactttcca catgacetta catggtgatg ccatccacgo ggaacggtgc actctatagg tgacacggac atttccaagt acggtgggag atacacccc attgaccatt atggctctt tacaacaacg tccgagcct acagtggagg cgaatctcgg gccgtggcgg gcagatggaa tgataagagt tgagcagtac ctgttccttt ttacatgatt acgttggctt aacctdccaa togggcaata cttatggtat gegtteeege gcgagcattc aatcgtcacc ccatggtata tcatctagtc tgccaaattc actcattgtc ctagcaacta aagcgaatgc ctacgccata atgetteaac cacttccagg gtacaatatg 781 901 961 1021 1081 1381 1441 1501 1561 1141 1201 1261 1321 1621 1681 1741 1801 1861 1921 2041 1981 2101 2161 2221 2281 2341 2521 2581 2401 2461 2641

> 10 15 20 25

30

ataatgatcc gtgccttcct ggtacctctc tatatatata tgaccaaagg ataagcagca attagccttg aaccaactta tcccggtaat agcaaggggc gttggtctgc aaaccttat tcccctaatg ggctcgtatg ttgattattg attgcatcgo ccatgattac ccgcggtggc gcttatcgat tatggagttc cccccccca ttatgcccag tccatatcat ccattgacgt gtatcatatq catogotatt tgactcacgg ccaaaatcaa cgcctggaga cggtaggcgt catacgaga taccgcctat ggcttggggc caaattcaca acttactcgt ggaaattatg ccctcccc aaatgaggaa ggggcagcac gggctctatg attgacccgg ttaacaaatg ctgtgtgttg cctctctctc agtgcctgtt gataataata cttgaaaaa gatccaagag cactgatgaa tgtcatatga ttatgcttcc acagctatga tggagctcca tcgatatcaa atacgttgta catgttgaca atagcccata cgcccaacga tagggacttt tacatcaagt ccgcctggca acgtattagt tgttttggca gatagoggtt cgcaaatggg accgtcagat accgatccag gtgacgtaag tactgtttt gcttagccta cttccatta caatactctg ttattattta acaagggaag tacgctttgg ctgttgtttg tctctcggta 9999t9999t ctctagagat tttcctaata gggatgcggt gtgctgaaga acaattactc ttatattatt tcagcaaatt gctttacact aaaaactgat ttggaatgaa tgaccutgag cacacaggaa gaacaaagc atacacatct ctgcaggaat tggccattgc acattaccgc tcattagttc cctggctgac gtaacgccaa cacttggcag ggtaaatggc aatgggcgtg caqtacatct cgtttagtga aatgggagtt gccccattga agacaccggg atgcatgcta ggctatatgc cgtgccaaga tgatggtata ggtgacgata ggggtcccat attatcttga ctacacaata aagatcagag ccactgtcc tctattctgg cacacatggt tgccagccat acatcacaac tetetetete teggtaceag aacgccattg aggcatgctg aaaataaaa ctatcattgg taattctcgt gttaaggtgg ggcacccag tcactaaagg atccccggg ctcatgtcca ataacaattt gattggctat aattacgggg aaatggcccg gtaaactgcc cgtcaatgac tectacttgg cattgacgtc tgttcccata gcagtacatc taacaactcc aagcagagct cctccataga geggatteee tttggctctt atgctatagg ctcccctatt tatctctatt tttacaggat tggctaataa gccttctagt aggtgccact taggtgtcat tetetetete atatttgaac aaaatttatt agacaatagc tctctctctc catcacttta gaagtgatgc gacggaatgt attgatgcct tggtaaggta aaatcattaa tcactcatta ttgtgagcgg caattaaccc gaactagtgg tegageatea tttatattgg aatagtaatc aacttacggt taatgacgta tatgggactt agtatttacg tgcggttttg caaaatgtcg aggtctatat gctgttttga ccctattga gtcttcaccc gcattggaac ggcacaccc ttattgacca ttgccacaac actctgtatt ccacttcctt gaagttttgc ctactagete agatcacttc gatctgctgt tgaccctgga attgtctgag aggattggga tetetetete tetetetete **Egecttttat** attaattatq cttagaaagt attttggtaa ccctatccaa aatacattac aagctttcct ttgtgtggaa gccaagcgcg gtgagttagc ggccgctcta accgctgacc aatatqtaca actagttatt cgcgttacat ttgacgtcaa caatgggtgg tcaagtacgc tacatgacct accatggtga ggatttccaa cgggacttc ctatacaccc gtacggtggg cgggaacggt acatggctct actgacacgg caccatccac agactctata ttattgacca 300I 3061 3121 3181 3241 3301 3361 3421 3481 3541 3601 3661 3721 3781 3841 3901 3961 4021 4141 4201 4261 4081 4321 4381 4441 4501 4561 4621 4681 4741 4801 4861 4921 4981 5041

20

25

30

35

ggagcttcca tacttgataa tcatctcgtg gggatctcca gtgccgcaca cgcacggctg gttgttgtat ggcagtgtag cagactaaca tattaaaatt tttggttctg gctgcaggag agcctctgga ggggctggag ctcctcagcc tgcacccgtg gcaaatgaat gataacattt ctctgggggc ggtgtcgtgg gtcctcagga caactggtac ccagacctac tgagcccaaa 699999*a* gacccutgag gtacaacagc tggcaaggag ggatgagctg catctccaaa cgacatcgcc tacagtgatg caggtggcag ctacacgcag tgagctgaca ctctggagat tggctccagc aggtgactac aacatagcgt cggtagcggc géteggeage caccaccagt agattgggct aggcagctga aacggtggag taatagetga atatttcgtt tegtgttege tctcctgtgc gatcaatcat cccaggtgca cagactatgo ctccagggaa cgttatatct cggggattat tggtcaccgt ccaagagcac aaccggtgac gcttgggcac ctgtcctaca tcgagaaaac acaagaaagt ctgaactcct tgatctcccg aggtcaagtt actggctgaa ccccatcccg tctatcccag agettteeta ggatcacctg agagattete tggaagatga gggaggagca agaccacgcc tggacaagag caggccaggc tgcacaacca gtttttatta ggctcttctc gctcatggtc caatgcccac atgagcgtgg aagaagatgo ccaccagaca ccgaaaggta cggtgctgtt gtcaccgtcg tctaaattta atatttttct tcccttagag gtccgcragg ggtggyacaa tcaaaaaaca cagggaaccc gcaccctcct acctttccgg tacttccccg ccctccagca accaaggtgg tgcccagcac ctgcaccagg tacaccetge tactgtacca gacaccctca gaagaccctg ccagccccca acaaagccgc gtcaaaggct gcagccggag aacaactaca aagctcaccg catgaggete ccagagccga caaacggcca cagcagaagt gggatccctg ggggcccagg gtgtctgaaa gcagcggcag tttctcaagg gaattggggc cgtgcccgca tccggacatg aggcacagca actecegttg **Bacacacaca** cttttctgca tgcgccagag 8668888558 gatgagetgg tacctatata agccgtatat cggcgtgcac gcctccagcg caaaattgat aagagatgat cttccccctg ggtcaaggac ggtgaccgtg gcccagcaac cctcaccgtc accacaggtg atgcccaccg aaaacccaag cgtgagccac taatgccaag caaagccctc cctctacago gacctgcctg ctccgtgatg gggtaaagcg gtccccagga ttattggtac acgaccetee gactatcagt caccatcccc gggtacgtgt ctggtcccat ggccagactt ggtagggtat aagacttaag gtcagaggta actcgttgct ttccatgggt gtgtggtgtt cagtttcggc ccgaggacac cgacgatgaa gcttggtaaa gaaacgcctg tecetecee gcccatcggt gtattaaaag tcaccatctc tgggctgcct tgaatcacaa aaactcacac tettecece tggtggtgga ccctgaccag ttagcaacgt tggaggtgca tggtcagcgt aggtetecaa agccccgaga aggtcagcct agagcaatgg geteettett tcttctcatg aaaaatatgt aggacagcaa cctgtctcc cggtgtcagt tggccacctt tatacaacaa catccgagcc taacagtgga tctgagcagt cgcgaatctc aggccgtggc acgcagatgg tctgataaga tcggggggag tgggtcggcc gactgttcct acttcttcqt taatattt gctttgtcaa ttcactttca agcctgaaag tccaccaagg aaaggcagat ggggagtta acagoggooc ctctacttcc aactcaggcg atctgcaacg tcttgtgaca tcagtcttcc gtcacatgcg gtggacggcg tacaagtgca gccaaagggc accaagaacc gtggagtggg gactccgacg gcattgccag tcagggacaa acgtaccgtg cadccaccct caggggaacg aagagcctct qtcatctatg 5401 5461 5581 5521 5641 5701 5821 5881 5941 6001 6121 6181 6241 6301 6061 6361 6421 6481 6541 1099 1999 6721 6841 7021 6781 6901 6961 7081 7141 7201 7261 7321 7381

2

30

35

ctacccggga accetectet ggagaccacc cctgacgctt gagcaccgtg ccttttgaa actgatgtgc atgtggcttc ctgagattt gtgccaggca cttcgcttc caattcgccc gaatggcgaa aagcaaaatg tgactgggaa cagctggcgt taaatcagct gatgatgtgc gaatagacco cgggatcata ggggattcat acctaaaaga tgagcitgaa aacgtggact aaaaaatgcc tgtttggata gtgacgaaat gaaggtcatc gccatcccaa accggttgca aattaaaaca actatctgca cctqccacaa ccatcatgct tttcctgtgt oggaagcata aagtgtaaag ctctqttccc taagtgactc aggcgggagt gctacctgag cgcacggaac cgcatgaagg aatgggtttc ctgcaggcca agggaagggc gtcactctga atatttcca attatctgtg tacaacgtcg ccctttcgc aaatttttgt actattaaag tctttatcgt tgcgcagcct cccactactc aaaatcatta catgcgaaaa gcccggtacc taaatcaaaa ccgtcaggaa agcttatgat ggctttttat acatcatttg tccaaatcca ggccaataac gatgactett ggattggcta gtcatagctg gggttaacat cagccaataa caaatgcact ggctattctt gaaagccaa tataccgatc caactgacag cagtggttat catagggagg ccctcggtca gtgtgtctca agccccgtca gcggccagca tgccaggtca ccgcggaggg tagtgaaaca tcaacatcta gcaccccttg tcgagggggg aattcgcgtt gccettcctc atccagttca ggtgtttgt gccgtcgttt gcagcacatc teceaacagt aaatccctta acaagagtcc accotctatc agggcgatgg ttaacttaat gatttttacc tatcgcaata tatttaccgc agctaaatct tcgcggaata tgagcttgag aaaccaataa gctaaaaagg aacagtaatg gcaccttgct atcccaccac cccatttagt cgtaatcatg gacccgtaat acatacgage acaccgtgct ataagtgatg taggtcagcc caaggetgee ggccacactg ggcagatagc caacaagtac aagctacagc agaatgttca atgactcctc tgcttcctct aaataaatc tccttccca catgagagaa ctgattaaca teceetegae gcgctcactg attttgttaa taatcgcctt gaaatcggca ccagtttgga cgatcgccct aatttattgc aattaacatt tggttatgca ttttatttga tcattatatt tttactccc gtaaaacgct aataacagca tccctgtaaa aggtaaagcg agatataaac cgttttttcq aaaagaccaa taatagcacc gcatcgttaa gagagattag attccacaca gaggagette aagecaacaa tggcctggaa acttcgcgcc aacaaagcaa agtcccacaa tggcccctgc ctgccattgc ttaagcacat ccacagcaag tactgctgcc gccctagatc teceattgee cgtattacgc aggcccgcac ccaataggcc ttacccaact aagcgttaat gagtgttgtt tgtatccacc aaaaaaggcc agggcgaaaa ggtcaacgag actcaatggc aatagatagg atcaagaggg ttgtctcctg taataataca aatgattata cctgtaatgc tcaacaaatg tcaccaataa accaaccttc gcagtactgc acgatttctg tactgagtgt tccgctcaca agggttaatt ctgaccgtcc tactgttact gccgtgacag acacceteca gagcagtgga gagaagacag 9999gaggaa tttgctactg agagggcct tcaagagtcg ccaacttggc tatagtgagt gattacatat acceacatta aaccctggcg aatagcgaag cattttttaa tggaaattgt agatagggtt tgacaagatg ccaacgtcaa agataaataa ctcttgggtt gatagcagga cagtgctcag ttaaaaactt gcttgccgat aactaagcac attggtagtg ttggtaaggc gacatcactc ataaatccga ccctttagtg gggaaaacta aggcttggaa attcatcatc gaataataa gaaattgtta 7801 7861 7921 8041 8101 7981 8161 8221 8281 8341 8461 8401 8581 8641 8521 8701 8761 8881 8821 8941 1906 9181 9001 9361 9121 9241 9301 9421 9481 9541 9601 1996

2

20

25

30

	9901			agctaactca	cattaattgc	gttgcgctca	ctgcccgctt	
	9961			tgccagctgc	attaatgaat	cggccaacgc	gcggggaga	
	10021		tattgggcgc	tcttccgctt	cctcgctcac	tgactcgctg	cgctcggtcg	
,	10081	treggetgeg	gcgagcggta	tcagctcact	caaaggcggt	aatacggtta	tccacagaat	
<u>ი</u>	10141	caggggataa	cgcaggaaag	aacatgtgag	caaaaggcca	gcaaaaggcc	aggaaccgta	
	10201			tttttccata	ggctccgccc	ccctgacgag	catcacaaaa	
	10261		aagtcagagg	tggcgaaacc	cgacaggact	ataaagatac	caggcgtttc	
	10321		ctccctcgtg	cgctctcctg	ttccgaccct	gccgcttacc	ggatacctgt	
•	10381	ccgcctttct	cccttcggga	agcgtggcgc	tttctcatag	ctcacgctgt	aggtatctca	
07	10441		ggtcgttcgc	tccaagctgg	gctgtgtgca	cgaaccccc	gttcagcccg	
	10501		cttatccggt	aactatcgtc	ttgagtccaa	cccggtaaga	cacgacttat	
	10561		agcagccact	ggtaacagga	ttagcagagc	gaggtatgta	ggcggtgcta	
	10621		gaagtggtgg	cctaactacg	gctacactag	aaggacagta	tttggtatct	
16	10681	gagatatgat	gaagccagtt	accttcggaa	aaagagttgg	tagctcttga	tccggcaac	
CT	10741	aaaccaccgc	tggtagcggt	ggttttttg	tttgcaagca	gcagattacg	cgcagaaaa	
	10801	aaggatctca	agaagatcct	ttgatctttt	ctacggggtc	tgacgctcag	tggaacgaaa	
	10861	actcacgtta	agggattttg	gtcatgagat	tatcaaaaag	gatetteace	tagatccttt	
	10921	taaattaaaa	atgaagttt	aaatcaatct	aaagtatata	tgagtaaact	tggtctgaca	
6	10981	gttaccaatg	cttaatcagt	gaggcaccta	tctcagcgat	ctgtctattt	cgttcatcca	
70	11041	tagttgcctg	actccccgtc	gtgtagataa	ctacgatacg	ggagggctta	ccatctggcc	
	11101	ccagtgctgc	aatgataccg	cgagacccac	gctcaccggc	tccagattta	tcagcaataa	
	11161	accagccagc	cggaagggcc	gagcgcagaa	gtggtcctgc	aactttatcc	geetecatee	
	11221	agtctattaa	ttgttgccgg	gaagctagag	taagtagttc	gccagttaat	agtttgcgca	
90	11281	acgttgttgc	cattgctaca	ggcatcgtgg	tgtcacgctc	gtcgtttggt	atggcttcat	
<b>C7</b>	11341	tcagctccgg	ttcccaacga	tcaaggcgag	ttacatgatc	cccatgttg	tgcaalaaag	
	11401	cggttagctc	cttcggtcct	ccgatcgttg	tcagaagtaa	gttggccgca	gtgttatcac	
	11461	tcatggttat	ggcagcactg	cataattctc	ttactgtcat	gccatccgta	agatgctttt	
	11521	ctgtgactgg	tgagtactca	accaagtcat	tctgagaata	gtgtatgcgg	cgaccgagtt	
(	11581	gctcttgccc	ggcgtcaata	cgggataata	ccgcgccaca	tagcagaact	ttaaaagtgc	
30	11641	tcatcattgg	aaaacgttct	tcggggcgaa	aactctcaag	gatcttaccg	ctgttgagat	
	11701	ccagttcgat	gtaacccact	cgtgcaccca	actgatcttc	agcatcttt	acttcacca	
	11761	gagtttatgg	gtgagcaaaa	acaggaaggc	aaaatgccgc	aaaaaaggga	ataagggcga	
	11821	cacggaaatg		atactcttcc	ttttcaata	ttattgaagc	atttatcagg	
3.0	11881	gttattgtct	catgagcgga	tacatatttg	aatgtattta	gaaaaataaa	caaatagggg	
2)	11941	ttccgcgcac	atttccccga	aaagtgccac				

#### **CLAIMS**

We Claim:

5

20

25

30

1. An isolated polynucleotide comprising at least one gene of interest and one or more pro nucleotide sequences, wherein the at least one gene of interest is operably-linked to a pro nucleotide sequence and each of the at least one gene of interest may be the same or different when more than one one gene of interest is present, wherein the at least one gene of interest encodes a protein useful in a vaccine.

- 2. The polynucleotide of claim 1, wherein a most 5' pro nucleotide sequence of the one or more pro nucleotide sequences is a part of a prepro nucleotide sequence.
- 3. The polynucleotide of claim 1, wherein two genes of interest and two pro nucleotide sequences are arranged in the following order: a prepro nucleotide sequence, a first gene of interest, a pro nucleotide sequence, and a second gene of interest.
  - 4. The polynucleotide of claim 3, wherein the prepro nucleotide sequence is a cecropin prepro nucleotide sequence and the pro nucleotide sequence is a cecropin pro sequence.
    - 5. The polynucleotide of claim 3, wherein the prepro nucleotide sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4 and the pro nucleotide sequence comprises a sequence shown in SEQ ID NO:1 or SEQ ID NO:2.
    - 6. The polynucleotide of claim 1, wherein a first gene of interest encodes for an antibody heavy chain and a second gene of interest encodes for an antibody light chain.
      - 7. A method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide comprising two or

5

more genes of interest, wherein each gene of interest encodes a part of the multimeric protein, each gene of interest is operably-linked to a pro nucleotide sequence, and each of the two or more genes of interest may be the same or different, wherein each of the two or more genes of interest encodes a protein useful in a vaccine.

- 8. The method of claim 7, wherein a most 5' pro nucleotide sequence of the two or more pro sequences is a part of a prepro nucleotide sequence.
- 9. The method of claim 8, wherein the polynucleotide comprises two genes of interest and two pro nucleotide sequences arranged in the following order: a prepro nucleotide sequence, a first gene of interest, a pro nucleotide sequence, and a second gene of interest.
- 15 10. The method of claim 7, wherein a first gene of interest encodes for an antibody heavy chain and a second gene of interest encodes for an antibody light chain.
- 11. A method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide comprising a cecropin prepro nucleotide sequence operably-linked to two or more genes of interest, each gene of interest encoding a part of the multimeric protein, wherein the multimeric protein is useful in a vaccine.
- 25 12. The method of claim 11, wherein a first gene of interest is an antibody heavy chain and a second gene of interest is an antibody light chain.
- 13. A method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide comprising two or more genes of interest, wherein each gene of interest encodes a part of the multimeric protein and wherein each gene of interest is operably linked to a gene encoding for a cleavage site, wherein the multimeric protein is useful in a vaccine.

14. The method of claim 13, wherein a transposon-based vector comprises the polynucleotide and further comprises a transposase gene operably linked to a first promoter and wherein;

5

- a) the first promoter comprises a modified Kozak sequence comprising ACCATG;
- b) the two or more genes of interest are each operably-linked to one or more additional promoters; and,

c) the two or more genes of interest and their operably-linked promoters are flanked by transposase insertion sequences recognized by a transposase encoded by the transposase gene.

15

- 15. The method of claim 13, wherein a transposon-based vector comprises the polynucleotide and further comprises a transposase gene operably linked to a first promoter and an avian optimized polyA sequence, and wherein;
- a) the two or more genes of interest are each operably-linked to one or more additional promoters; and,
- b) the two or more genes of interest and their operably-linked promoters are flanked by transposase insertion sequences recognized by a transposase encoded by the transposase gene.
- 16. An animal comprising the isolated polynucleotide of Claim 1.
- 17. The animal of Claim 16, wherein the animal is a bird.

25

20

- 18. An egg produced by the animal of Claim 16.
- 19. The egg of Claim 18, wherein the egg comprises a multimeric protein encoded by the isolated polynucleotide.

30

- 20. The animal of Claim 16, wherein the animal is a mammal.
- 21. Milk produced by the mammal of Claim 20.

i.

22. The milk of Claim 21, wherein the milk comprises a multimeric protein encoded by the isolated polynucleotide.

- 5 23. A method of producing a multimeric protein comprising:
  - a) administering to an egg-laying animal a composition comprising the polynucleotide of Claim 1; and,
  - b) permitting the one or more genes of interest to be expressed into the multimeric protein, wherein the multimeric protein is useful in a vaccine.
  - 24. The method of Claim 23, further comprising

10

15

20

25

30

- a) collecting an egg from the egg-laying animal;
- b) harvesting egg white comprising the multimeric protein; and,
- c) purifying the multimeric protein.
- 25. The method of Claim 23, wherein the egg-laying animal is a bird.
- 26. A method of producing a multimeric protein comprising:
- a) administering to an intramammary duct system of a mammal a composition comprising the polynucleotide of Claim 1, and,
  - b) permitting the one or more genes of interest to be expressed into the multimeric protein, wherein the multimeric protein is useful in a vaccine.

27. The method of Claim 26, further comprising

- a) collecting milk from the mammal, wherein the milk comprises the multimeric protein; and
  - b) purifying the multimeric protein.
- 28. A method of vaccinating an individual against a disease comprising administering to the individual a composition comprising a protein, a multimeric protein or an antibody or a combination thereof, wherein the

composition is produced in the individual or in another individual made transgenic using transposon-based vectors.

29. Use of a composition comprising a protein, a multimeric protein or an antibody or a combination thereof, wherein the composition is produced in an individual made transgenic using transposon-based vectors, in the preparation of a medicament for use in a vaccine.

5

15

- 30. The polynucleotide of any of the preceding claims, wherein the prepro nucleotide sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4.
  - 31. The polynucleotide of any of the preceding claims, wherein the two or more pro nucleotide sequences each comprise a sequence shown in SEQ ID NO:1 or SEQ ID NO:2.
  - 32. The polynucleotide of any of the preceding claims, wherein the prepro nucleotide sequence is a cecropin prepro nucleotide sequence.
- 20 33. The polynucleotide of any of the preceding claims, wherein the pro nucleotide sequence is a cecropin pro nucleotide sequence.

**IGURE 1** 

Prom prepro Gene of Interest pro Gene of interest polyA

1/5

'IGURE 2

IS Oval Prom prepro Heavy chain pro Light chain polyA IS	ſ	
Oval Prom prepro Heavy chain pro		IS
Oval Prom prepro Heavy chain pro		
Oval Prom prepro Heavy chain pro		polyA
Oval Prom prepro Heavy chain pro		Light chain
Oval Prom prepro Heavy		pro
Oval Prom F		Heavy chain
Oval		prepro
IS		Oval Prom
		IS

polyA

Gene of Interest

IGURE 3

Prom | Cecropin prepro | Gene of Interest | CS

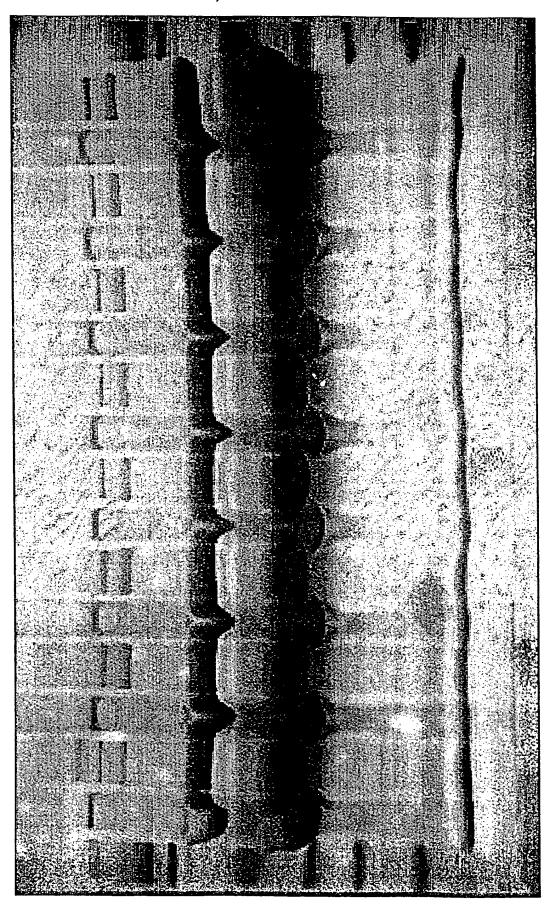
4/5

GURE 4

IS Prom SS Gene of Interest CS Gene of Interest polyA IS

4/5





# This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

#### **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

□ BLACK BORDERS
□ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
□ FADED TEXT OR DRAWING
□ BLURRED OR ILLEGIBLE TEXT OR DRAWING
□ SKEWED/SLANTED IMAGES
□ COLOR OR BLACK AND WHITE PHOTOGRAPHS
□ GRAY SCALE DOCUMENTS
□ LINES OR MARKS ON ORIGINAL DOCUMENT
□ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

#### IMAGES ARE BEST AVAILABLE COPY.

☐ OTHER:

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.